

Qy	5879	ctcgcaagccccagcccgctccgggtctgagcccgctcctccccagcaagggtctg	5938
Db	5881	ctcgcaagccccagcccgctccgggtctgagcccgctcctccccagcaagggtctg	5940
Qy	5939	gagcccgccctagtgcctctctctgtgagccagccacatcgcccgacccctcg	5998
Db	5941	gagcccgccctagtgcctctctctgtgagccagccacatcgcccgacccctcg	6000
Qy	5999	aagaacctcgacctcaccagccagcccgagcccgcccgcccgccacctgctcgccctg	6058
Db	6001	aagaacctcgacctcaccagccagcccgagcccgcccgcccgccacctgctcgccctg	6060
Qy	6059	gacccgcacccggaaaaagactcaaatgataaaccttttccatccaggaactggaaactcgt	6118
Db	6061	gacccgcacccggaaaaagactcaaatgataaaccttttccatccaggaactggaaactcgt	6120
Qy	6119	tcctgggttacacggcgagcgtcacagcccgaaaggggtggagccgtcagccctgtg	6178
Db	6121	tcctgggttacacggcgagcgtcacagcccgaaaggggtggagccgtcagccctgtg	6180
Qy	6179	agtcaccagttgacccacgacaagggttccccagacctgggaagagctcgacaag	6238
Db	6181	agtcaccagttgacccacgacaagggttccccagacctgggaagagctcgacaag	6240
Qy	6239	agccacctggaggggagctgcggcccaagcagccagcccgctgaagacttgcggggag	6298
Db	6241	agccacctggaggggagctgcggcccaagcagccagcccgctgaagacttgcggggag	6300
Qy	6299	gcgcaccactcccacactcgccgctgctgaagcagccctcgtccagcccgctg	6358
Db	6301	gcgcaccactcccacactcgccgctgctgaagcagccctcgtccagcccgctg	6360
Qy	6359	ctccagacgcgccaggggtcaaaagtccacagcgggtgggtcaccttggccagcacatc	6418
Db	6361	ctccagacgcgccaggggtcaaaagtcaccagcgggtgggtcaccttggccagcacatc	6420
Qy	6419	agtgaagtatacacagagactacaccgggcaacaccacacagcagctcagccacctg	6478
Db	6421	agtgaagtatacacagagactacaccgggcaacaccacacagcagctcagccacctg	6480
Qy	6479	cccgccccctctactccttcctggggcagctgcccgctctggacctccgcgcgcca	6538
Db	6481	cccgccccctctactccttcctggggcagctgcccgctctggacctccgcgcgcca	6540
Qy	6539	ccagtgacctctacctcccgcccccgacatatggtccccggccgctgggtccccccac	6598
Db	6541	ccagtgacctctacctcccgcccccgacatatggtccccggccgctgggtccccccac	6600
Qy	6599	agcgaagggggcaagagggtctccagagccaaacaagacgttcgttgggttgggtgag	6658
Db	6601	agcgaagggggcaagagggtctccagagccaaacaagacgttcgttgggttgggtgag	6660
Qy	6659	gacggtatgaacctgtccccaccgagggtcagcagcagcagggcactcccgagt	6718
Db	6661	gacggtatgaacctgtccccaccgagggtcagcagcagcagggcactcccgagt	6720
Qy	6719	gctgttaccgcgtgtctaccgggatgggaacagcggagcccgaggatgggtctcc	6778
Db	6721	gctgttaccgcgtgtgtaccgggatgggaacagcggagcccgaggatgggtctcc	6780
Qy	6779	aagtctccaggcaacaccgacgcgcgaccttcttcagcaagctgaccagagacaac	6838
Db	6781	aagtctccaggcaacaccgacgcgcgaccttcttcagcaagctgaccagagacaac	6840
Qy	6839	tcgccaatggtcaagtccaagaagcaagagatccaacaagaagctgaacacccaacccg	6898
Db	6841	tcgccaatggtcaagtccaagaagcaagagatccaacaagaagctgaacacccaacccg	6900
Qy	6899	aatgagcctgaatacataatcatcgccagcctgggacgagatcttcaatatgtccgcctc	6958
Db	6901	aatgagcctgaaatacataatcatcgccagcctgggacgagatcttcaatatgtccgcctc	6960
Qy	6959	accggaacaggccttatgacctatagaagccagcggtgtcaggaacaatgcagacccaac	7018

Qy	8407	tgtctg-tcacgcccgctgtgtctccct-ccctccctctcttgggcagaataatcg	8464
Db	805	tgtggtcaagcccgctgtgtctccctccctcccttcttgggcagaataatcg	864
Qy	8465	atgcgtattctgtgcccgcatttgcgcagggtgtgtattctgtcatttacacagctc	8524
Db	865	atgcgtattctgtggcgcacatctgcgcagggtgtgtattctgtcatttacacagctc	924
Qy	8525	gtctcaattaaaaagcggaattatctccaaa	8556
Db	925	gttctcaattaaaaagcggaattatctccaaa	956

RESULTS

RESULTS
22299744

AAA98144
ID AAA98144 standard: cDNA: 956 bp.

AC AAA98144:

26-JAN-2001 (first entry)

DE Human proliferation-associated protein EST encoded cDNA #8.

KW proliferation-associated protein; differentiation-associated protein;
 KW ZST; expressed sequence tag; endothelial cell; anti-angiogenic;
 KW angiogenic; antiproliferative; anti-arthritis; antifibrotic; hepatotropic;
 KW antiatherosclerotic; nephroprotective; antidiabetic; ophthalmological;
 KW immunosuppressive; neuroprotective; gene therapy; psoriasis; arthritis;
 KW hemangioma; diabetic retinopathy; glomerulonephritis; liver cirrhosis;
 KW transplant rejection; arteriosclerosis; nervous tissue injury; human; ss.
 XX
 XX Homo sapiens.
 OS

PN WO200053734-A2.

14-SEP-2000.

08-MAR-2000; 2000WO-EP02005.

09-MAR-1999; 99DE-1011684.

PR 01-OCT-1999; 99DE-1048679.

PA (SCHED) SCHERING AG.

PI Thierauch K, Glienke J, Hinzmann B, Pilarsky C;

DR WPI: 2000-572267/53.

Nucleic acid sequences from human endothelial cells, useful for gene therapy of angiogenesis and for identifying antiangiogenic agents -

PS Claim 1a: Page 95; 115pp: German.

This invention describes novel nucleic acid sequence (I) obtained from an endothelial cell comprising any of 59 sequences (defined and given in the specification), its allelic variant, or complement. The products of the invention have anti-angiogenic, angiogenic, antiproliferative, anti-arthritic, antifibrotic, antiarteriosclerotic, nephrotropic, antidiabetic, ophthalmological, hepatotropic, immunosuppressive and neuroprotective activity. (I) are implicated in differentiation and proliferation of endothelial cells. (I) are used for expression of polypeptides (IV) that are useful for identifying agents (A) for treatment of angiogenic disease, to express (IV) from gene therapy vectors for treatment of such diseases and as antisense reagents. (IV) are also used to raise specific antibodies. Angiogenic diseases that may be treated include psoriasis, arthritis, hemangioma, diabetic retinopathy, glomerulonephritis, transplant rejection, liver cirrhosis, arteriosclerosis and injuries to nervous tissue. AAA98102-A98161 represent proliferation-associated and differentiation-associated EST (expressed sequence tag) derived sequences which are described in the method of the invention.

XX

SQ Sequence 956 BP; 192 A; 288 C; 287 G; 189 T; 0 other;

Query Match 10.5%; Score 898.4; DB 21; Length 956;
Best Local Similarity 99.1%; Pred. No. 2e-144;
Matches 924: Conservative 0; Mismatches 6; Indels 2;

QY	7627	tgcaggagcggcggtgctccgactcccccaacaaaggaaagcccctgagtccgcgtg	7686
Db	25	tgcaggagcaggcggtgcgactgcccaaccaaggaaagcccctgagtccgcgtg	84
QY	7687	cgcctccatccattctgcgtccagagcggcatccttgcgtgtctaaaagccttaataa	7746
Db	85	cgcctccatccattctgcgcgcagcgcgcgatccttgcgtgtctaaaagccttaataa	144
QY	7747	gaactccgcgcccggtgcgttcgtgcagacctactcaggggatgtttacctggtgctc	7806
Db	145	gaactccgcgcccggtgcgttcgtgcagacctactcaggggatgtttacctggtgctc	204
QY	7807	ggaaaggggggaaggcggggaggggaggggcagcgcagcgtgtgagccaccacacag	7866
Db	205	ggaaaggggggaaggcggggaggggaggggcagcgcagcgtgtgagccaccacacag	264
QY	7867	gcggccagggcgccagggaccaaagcaggatgaccacgcacotccaagcctgcctgc	7926
Db	265	gcggccagggcgccagggaccaaagcaggatgaccacgcacotccaagcctgcctgc	324
QY	7927	ccccgaatgcatttggacaacaaagtctaactgagctcgcgcagccccgcctccctcc	7986
Db	325	ccccgaatgcatttggacaacaaagtctaactgagctcgcgcagccccgcctccctcc	384
QY	7987	gcctccatcccgcttagcgtcttgagacagatggacgcagggcctgtccagccccagtg	8046
Db	385	gcctccatcccgcttagcgtcttgagacagatggacgcagggcctgtccagccccagtg	444
QY	8047	cgcctgtccgctccccacagactccccagcagcagcagcagcagcagcagcagcagc	8106
Db	445	cgcctgtccgctccccacagactccccagcagcagcagcagcagcagcagcagcagc	504
QY	8107	ccaggctgggcggacaaaaggccaggtgcgcgcctggggggaacggatgctccagggactg	8166
Db	505	ccaggctgggcggacaaaaggccaggtgcgcgcctggggggaacggatgctccagggactg	564
QY	8167	gaactgtttttcacacatcgtgtccgcagcgcgtgggaagaaaggcagatgtaaatgat	8226
Db	565	gaactgttttttcacacatcgtgtccgcagcgcgtgggaagaaaggcagatgtaaatgat	624
QY	8227	gtgttgttttacagggtatatttttgataccttcaatgaattaactcagatgttttacgc	8286
Db	625	gtgttgttttacagggtatatttttgataccttcaatgaattaactcagatgttttacgc	684
QY	8287	aaggagaagacttaccagtatatttactgctgtgcttttgccttctctcttaccgttcaag	8346
Db	685	aaggagaagacttaccagtatatttactgctgtgcttttgccttctctcttaccgttcaag	744
QY	8347	aggcgtgtcaggccgcagtcggtgacccccatcactcgcaggaccaggggcgggggac	8406
Db	745	aggcgtgtcaggccgcagtcggtgacccccatcactcgcaggaccaggggcgggggac	804
QY	8407	tgcctcg-tcaaccccccgctgtctccct-ccctccctctcttgggcagaatgaattcg	8464
Db	805	tgcctcgctaacgcccccgctgtgtctccctccctccctctcttgggcagaatgaattcg	864
QY	8465	atgcgtattctgtggccgcatttgcgcagggttggtattctgtctatttacacagctc	8524
Db	865	atgcgtattctgtggccgcattctgcgcagggttggtattctgtctatttacacagctc	924
QY	8525	gttctaattaaaaagcgaattatactccaaa-8556	
Db	925	gttctaattaaaaagcgaattatactccaaa-956	

RESULTS

QY 1620 acgacaagaagactctcaagagagaagacagacacacctcaggaggaacagcag 1679
|||||
Db 61 acgacaagaagactctcaagagagaagacagacacacctcaggaggaacagcag 120
|||||
QY 1680 agaagagagctgtgctcccaaaagccgcaaaactgccaaacagccagggagagacgcaag 1739
|||||
Db 121 agaagagagctgtgctcccaaaagccgcaaaactgccaaacagccagggagagacgcaag 180
|||||
QY 1740 gccgatcacccgctcaatgctaatgagccaaacagcagcagcagccatccccccagc 1799
|||||
Db 181 gccgatcacccgctcaatgctaatgagccaaacagcagcagcagccatccccccagc 240
|||||
QY 1800 agagcgcgagctgctcccaatggagctgaatgagagttctcgtggacagagaagaaaa 1859
|||||
Db 241 agagcgcgagctgctcccaatggagctgaatgagagttctcgtggacagagaagaaaa 300
|||||
QY 1860 tggaaacagcca 1871
|||||
Db 301 tggaaacagcca 312
|||||
RESULT 10
ID AAT25693 standard; cDNA to mRNA; 226 BP.
XX AC AAT25693;
XX AC
DT 10-OCT-1996 (first entry)
XX XX
DE Human gene signature HUMGS07893.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUBO/) OKUBO K.
XX
XX Matsubara K, Okubo K;
PI
XX
XX
DR WPI; 1995-206931/27.
XX
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
XX Claim 1; Page 1908; 2245pp; Japanese.
XX
XX A single-stranded DNA (or its complementary strand or the corresp.
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences
XX given in AAT19001-T26837 and which is able to hybridise to part of
XX the human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
XX sequences were obtained from 3'-directed cDNA libraries prepared
XX from various human tissues; synthesis of cDNA was initiated from the
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
XX untranslated sequence is unique to a particular mRNA species, almost
XX all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
XX is constructed so as to reflect accurately the relative abundance of
XX different mRNAs in the particular tissue from which it was derived.
XX The appearance frequency of a given GS in a cDNA library can be
XX determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.

XX
SQ Sequence 226 BP; 43 A; 65 C; 61 G; 56 T; 1 other;

Query Match 2.4%; Score 206.2; DB 16; Length 226;
Best Local Similarity 97.8%; Pred. No. 1.2e-26;
Matches 219; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 8326 gatctctgtaccgttcaagagggcgtgtgcagggccagcagtcgtagcccatcactcg 8385
|||||

Db 1 gatctctgtaccgttcaagagggcgtgtgcagggccagcagtcgtagcccatcactcg 60
|||||

QY 8386 caggaccaagggcgaggactgctcg-tcacgccccgctgtgtctcctccctccctt 8444
|||||

Db 61 caggaccaagggcgaggactgctcg-tcacgccccgctgtgtctcctccctccctt 120
|||||

QY 8445 ccttgggcagaatgaattcgcgtattctgtggccgccatttgcaggggtggtgta 8504
|||||

Db 121 ccttgggcagaatgaattcgcgtattctgtggccgccatttgcaggggtggtgta 180
|||||

QY 8505 ttctgtcatttacacacgctcgttctaataaaaaagcgaattata 8548
|||||

Db 181 ttctgtcatttacacacgctcgttctaataaaaaagcgaattata 224
|||||

RESULT 11

AAZ16912

ID AAZ16912 standard; cDNA; 789 BP.

XX AC AAZ16912;

XX AC

XX XX

DT 12-OCT-1999 (first entry)

XX XX

DE Human gene expression product cDNA sequence SEQ ID NO:4382.

XX Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.

OS

XX

XX

PN WO938972-A2.

XX

XX

PD 05-AUG-1999.

XX

XX

PF 28-JAN-1999; 99WO-US01619.

XX

XX

PR 03-APR-1998; 98US-0080666.

PR

PR 28-JAN-1998; 98US-0072910.

PR

PR 24-FEB-1998; 98US-0075954.

PR

PR 31-MAR-1998; 98US-0080114.

PR

PR 03-APR-1998; 98US-0080515.

XX

XX

XX (CHIR) CHIRON CORP.

PA

PA (HYSE-) HYSEQ INC.

XX

XX

PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Glese K, Innis MA;

PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX

XX WPI; 1999-494092/41.

XX

XX Novel human genes and their expression products which are

PT differentially expressed in different cell types

XX

PS Claim 1; Page 2075; 2479pp; English.

XX

CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ1779. Also described is

RESULT	13
AAT31931	
ID	AAT31931 standard; DNA; 1922 BP.
AC	XX
AC	AAT31931;
XX	
DT	27-SEP-1996 (first entry)
XX	
XX	Retinoid X receptor interacting protein RIP13 DNA.
XX	
KW	Retinoid X receptor interacting protein; RXR; RIP; RIP13; ss.
XX	
OS	Mus sp.
XX	
Key	Location/Qualifiers
FH	2..1861
FT	/ *tag= a
FT	
XX	
PX	WO9621677-A1.
XX	
PD	18-JUL-1996.
XX	
PF	08-DEC-1995; 95WO-US16311.
XX	
PR	13-JAN-1995; 95US-0372652.
XX	
PA	(GEHO) GEN HOSPITAL CORP.
XX	
PI	Choi H, Moore D, Seol W;
XX	
DR	WPI; 1996-342241/34.
DR	P-PSDB; AAR99738.
XX	
PT	Retinoid X receptor (RXR) interacting protein (RIP) - useful to
PT	modulate or mediate RXR function, anti-RIP antibodies can be used to
PT	determine RIP subcellular distribution patterns
XX	
PS	Claim 8; Page 56-57; 90pp; English.
XX	
CC	A CDNA clone (AAT31931) codes for retinoid X receptor interacting
CC	protein RIPI3 (AAR99738), which is a candidate transcriptional
CC	co-activator. It was isolated from a mouse liver cDNA library.
CC	The CDNA can be used for the recombinant prodn. of RIPI3 in
CC	transformed host cells.
XX	
SQ	Sequence 1922 BP; 545 A; 508 C; 451 G; 418 T; 0 other;
Query Match	1.7%; Score 142; DB 17; Length 1922;
Best Local Similarity	52.5%; Pred. No. 1.4e-15;
Matches	435; Conservative 0; Mismatches 375; Indels 18; Gaps 5;
QY	6730 gctgctgtaccgggatgggaacagcaggccacagcagatgggtccaagtctccagg 6789
DG	1039 gctcttgctcacaggaggagtgagacctgctgagcaaggagtgattctcgatcacagg 1098
QY	6790 caacaccagccgcgcagcttcttcagcaagctgaccgagacaactcgcgcattgt 6849
DG	1099 aagtataagctacttgctttcttcttccaccaagt---tgaaagcacatcacccattgt 1155
QY	6850 caagtccaagaagcaagatcaacaagaagctgaacccccacacccgaatgagcttga 6909
DG	1156 taaatcaagaacacagaaattttctgaagttgaactcttctgtgaggtgactctga 1215
QY	6910 atacaatatcagcagcctgggacggagatcttcaatatgcccgcatcaccggaacagg 6969
DG	1216 tatggcagctgctcagcaggaacagagatcttcaattcgcagcagttaccacatcagg 1275
QY	6970 ccattagacctatagaagccagcggtgcaggaaacatgccagacccaacatggggctga 7029
DG	1276 tgcaagtgcagtaagaagccattcttctgtgatcccgcag---taaccttgggtctaga 1332
QY	7030 ggccataattagaagcactcatcaggttaaat---atgaccagtggaagagatccccgcc 7086

XX PS Claim 9; Page 1037; 1046pp; English.

XX CC The present sequence is one of 3351 sequences in a library of human

CC CC polynucleotides. The library is used to detect differentially expressed

CC CC genes correlated with a cancerous state of a mammalian cell and can

CC CC detect colon, prostate, breast and lung cancer. The library can be used

CC CC to produce probes for detection of mRNA and to produce additional copies

CC CC of the polynucleotides. The probes can be used for chromosome mapping of

CC CC the polynucleotide and for detection of transcription levels. Ribozymes

CC CC or antisense oligonucleotides can be generated. The polynucleotides and

CC CC their gene products are used as genetic or biochemical markers (e.g. in

CC CC blood or tissues) that will detect the earliest changes along the

CC CC carcinogenesis pathway and/or monitor the efficacy of therapies and

CC CC preventive interventions. The polynucleotides, polypeptides and

CC CC antibodies against them can be used in pharmaceutical compositions to

CC CC treat the cancers and proliferative disorders such as neoplasia,

CC CC dysplasia and hyperplasia.

XX SQ Sequence 342 BP; 135 A; 57 C; 81 G; 68 T; 1 other;

Query Match 1.5%; Score 130.8; DB 22; Length 342;

Best Local Similarity 67.3%; Pred. No. 8.6e-14;

Matches 214; Conservative 0; Mismatches 77; Indels 27; Gaps 1;

Qy 660 agtcgaagcaccgcgcgctgtgagatcattctacgacgagaaacggagagctgaag 719

Db 24 agggaaacacgcgcgattgtccaaattatttatgatgagaatcggaataagcagaag 83

Qy 720 ctgcacatcgattctgaaagcctgtggcccccagctgagctgcctgtacaaccagc 779

Db 84 aagctcataaaattttgaaagctgtggcccaaaagtgtgaactgcactgtataaccagc 143

Qy 780 cctccgacacccgcgactatcagaacacaa----- 813

Db 144 catcagataccaaggtgcaccatgagacatacaactgagctgagctgaagcgcatga 203

Qy 814 -aataaacacgcgctgcggaagactaattctgtacttcaagagggaggaatcacgctc 872

Db 204 tgaataacacgcgctgcggaagactaattctgtacttcaagagggaggaatcacgctc 263

Qy 873 ggaataacatgaagcagagcttctgcagcgtatgaccagctatgagccttggaaa 932

Db 264 gaaataacagggaaacaaatactgccagcgttatgatcagctcatgagcgtgggaga 323

Qy 933 aaaagtgtagcgcatcg 950

Db 324 aaaaagtggacagaatag 341

RESULT 15

AA53491/c

ID AA53491 standard; DNA; 114955 BP.

AC AA53491;

XX 05-JUL-1999 (first entry)

XX Human adenosine A1 receptor antisense oligonucleotide fragment.

XX Antisense oligonucleotide; multiple target; antisense treatment;

KW impaired respiration; inflammation; lung disease;

KW pulmonary vasoconstriction; inflammation; allergic rhinitis;

KW acute asthma; allergy; asthma; impeded respiration;

KW respiratory distress syndrome; pain; cystic fibrosis;

KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;

KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;

KW colon cancer; breast cancer; lung cancer; pancreatic cancer;

KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;

XX prostate cancer; ss.

XX Synthetic.

XX WO9913886-A1.

XX 25-MAR-1999.

XX 17-SEP-1998; 98WO-US19419.

XX 09-JUN-1998; 98US-0093972.

XX 17-SEP-1997; 97US-0059160.

XX (UYEC-) UNIV EAST CAROLINA.

XX Nyce JW;

XX WPI; 1999-229400/19.

XX New antisense oligonucleotides used in treatment of, e.g. pulmonary

PT vasoconstriction

XX Disclosure; Page 37; 120pp; English.

XX The specification describes antisense oligonucleotides (AA52869-X55271)

CC directed against at least 2 mRNAs selected from target genes, coding and

CC non-coding regions of RNAs corresponding to target genes, gene

CC initiation codons, genomic flanking regions, intron-exon borders, the

CC 5'-end, the 3'-end and the juxta-section between coding and non-coding

CC regions and all segments of RNAs encoding proteins associated with one

CC or more diseases, conditions or mixtures. The antisense oligonucleotides

CC may be derived from sequences AA55272-74. These multiple target

CC oligonucleotides (specifically AA55180-271) can be used for the

CC antisense treatment of diseases and conditions. Typical diseases and

CC conditions are those associated with impaired respiration and

CC inflammation, including lung diseases, pulmonary vasoconstriction,

CC inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded

CC respiration, respiratory distress syndrome, pain, cystic fibrosis,

CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic

CC obstructive pulmonary disease (COPD), and cancers such as leukemias,

CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,

CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,

CC hepatic metastases, as well as all types of cancers which may metastasize

CC or have metastasized to the lungs, including breast and prostate cancer.

XX SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

Query Match 1.4%; Score 118.4; DB 20; Length 114955;

Best Local Similarity 29.7%; Pred. No. 3.1e-11;

Matches 972; Conservative 280; Mismatches 1987; Indels 32; Gaps 6;

Qy 2343 gccacaccacccacccacgcgagagacatcccgggcccccattgagccaccccgccctcg 2402

Db 111316 GCTGGCCACCGCCGCCCAAGTGGAGTGGCTGGGCGGCCCCACVCVAAACGGGACCG 111257

Qy 2403 aagccacccgagccctacgccccacacagcaccacccatcgccctctgcacctcctcg 2462

Db 111256 CGCGGACAVGCVCAAVGVCACCNHNNHNNNSGCCGAGAGCGCGCGTGCCTGCNNH 111197

Qy 2463 tggctcccaaggaggaaggaaggaagacccagcagcgcgcacccagtgagagagggg 2522

Db 111196 NNNNSCCGACGCGCGGAGGAGCGCCCNHNNHNNNSGCCGCGCGGCGGAGAGACNN 111137

Qy 2523 aggcagagaaccccccgcgctgagagctggcagtggaacacagggaggaagccagagc 2582

Db 111136 HNNNSGCCGACCGCCVCCVCCGCCGCCGCGGNNHNNHNNNSGCCGCGGACCGCGCGC 111077

Qy 2583 ccgtcaagagcagtgacaggaaggaagcagagagggggccgaggaagcagcgcg 2642

Db 111076 CVCGCGCNHNNHNNNSCCGAGCGGAGGAGCAGCNHNNHNNNSGCCGAGAGCGCGCGC 111017

Qy 2643 aggcgctgagccacgcgagggggcgctcaagcagaggaaggaagagggcgagcg 2702

Db 111016 CNHNNNSCCGCGGAGCGCGGAGGAGCGCCCNHNNHNNNSGCCGCGGCGGCGGAGGCGCGCGC 110957

[illegible]

Db	109885	VGCGCGACCNHNHNNSGVCVVGCGCCGACCNHNHNNSAGCVCVVGCGCCGACCNHNHNNSG	109826
Qy	3837	ccctatgaggggtggcatgtctgtgaccagtgctctcaagggagcggcgaagaagcagctcag	3896
Db	109825	AGCVCVVGCGCGACCNHNHNNSCGAGCVCVVGCGCGACCNHNHNNSACGAGCVCVVGCGC	109766
Qy	3897	gaccccccatgagacgccccgcgcccaagcgacctatgacatgatggaggggccgcgtgg	3956
Db	109765	GACCNHNHNNSGACGAGCVCVVGCGCGACCNHNHNNSCGACGAGCVCVVGCGCGACCNH	109706
Qy	3957	gcagagccatctctcagcagcatcgaaggtctcatgggcgctgcacatccgcgcggagc	4016
Db	109705	NNNSACGACGAGCVCVVGCGCGACCNHNHNNSACGACGAGCVCVVGCGCGACCNHNHN	109646
Qy	4017	gacacgccccccacactcaagagcagcacacatccgcggctccatcacacaaggga	4076
Db	109645	SCCAGCAGCAGCVCVVGCGCGACCNHNHNNSGCCACGACGAGCVCVVGCGCGACCNHNHN	109586
Qy	4077	tcctcgtctcactgtaggagcacagaggaactacgtcgctgggaggccaaagctcctaa	4136
Db	109585	NSGGCGCGVGGCCACCCCNHNHNNSGGGCCCVAGNHNHNNSGGGCCCVAGNHNHN	109526
Qy	4137	agcgggagggcacgctctccgccccccacgcgcctcactcactgacgcagagcctaca	4196
Db	109525	SGGGGCCCVAGCNHNHNNSGGGCCCVAGCGCNHNHNNSGGGCCCVAGCGANNHNHN	109466
Qy	4197	gcagggccctggccccctgaagctgaagcggcccccatgaagcctggctggccacgtga	4256
Db	109465	SGGGGCCCVAGCGAGNHNHNNSGGGCCCVAGCGAGCNHNHNNSGGGCCCVAGCGA	109406
Qy	4257	-----aggagggcgccgtccatccatgatacccgcgagagagctcgcg	4304
Db	109405	GCCNHNHNNSGGGCCCVAGCGAGCCANNHNHNNSGGGCCCVAGCGAGCCANNHNHN	109346
Qy	4305	acagcccgagctgccctggccccgcgcgcgtcaaggagggtccatacagcagggca	4364
Db	109345	SGGGGCCCVAGCGAGCCAGNHNHNNSGGGCCCVAGCGAGCCAGCNHNHNNSGGGG	109286
Qy	4365	ccccctcaagtacgacacggcgctccaccactgctccaaaaagcagcagctacgt	4424
Db	109285	CCCVAGCGAGCCAGCGCNHNHNNSGGGCCCVAGCGAGCCAGCGCNHNHNNSGGGGC	109226
Qy	4425	ccctcatcggcagccccgcgcgagcttcccaaccgtgcaccgcgtggatgtgatggcg	4484
Db	109225	CCVAGCGAGCCAGCGCCCNHNHNNSGGGCCCVAGCGAGCCAGCGCCGNHNHNNSGG	109166
Qy	4485	acgccccggcactggaactgtctgtacgagagagcgtgaagagccgcgcaggagccg	4544
Db	109165	GGCCVAGCGAGCAGCAGCGCGCCGNHNHNNSGGGCCCVAGCGAGCAGCGCGCCCN	109106
Qy	4545	ccacagctcggggggtccatcgcgcggcgcgcgcgttcattgtgctgagctgggta	4604
Db	109105	HNNNSGGGCCCVAGCGAGCCAGCGCGCCGNHNHNNSGGGCCCVAGCGAGCCAGCGC	109046
Qy	4605	agccgcgcagagccccctgacctatgagcacacgggcacaccttgcgcgcacctcc	4664
Db	109045	CGCGCGCNHNHNNSGGGCCCVAGCGAGCAGCGCGCGCCGNHNHNNSGGGGCCC	108986
Qy	4665	cacgaggttcgcgcgtgacctcgggagccccgcgcgcgtgcaggaggcagccttt	4724
Db	108985	VAGCGAGCAGCGCGCGCGCCCNHNHNNSGGGCCCVAGCGAGCAGCGCGCCCGC	108926
Qy	4725	cgtccagcaaggcatccagagcagaaagctgacgttcagcgcctctgtgagatcgccaagt	4784
Db	108925	CCCNHNHNNSGGGCCCVAGCGAGCAGCGCGCGCGCCCNHNHNNSGGGGCCVAG	108866
Qy	4785	ccccgcagcacacgtgcccgagcacaccaccaccccctctcgccctatgacacctgc	4844
Db	108865	GCGAGCAGCGCGCGCGCCCNHNHNNSGGGGCCCVAGCGAGCAGCGCGCGCGCGC	108806
Qy	4845	ttcggggcgtagtggcgtagacctgtatcgagccacataccccctggcctctcgaccccc	4904

Db 108805 CCCCCGNNNNNNNSCGCCCCCGCCGNNNNNNNSGCCGCCGCCGNNNNNNNSCGCCCCGCC 108746
Qy 4905 cctccatacccgcgccatccctctgagcagcgctgctactactgctgccccgacacc 4964
Db 108745 GCNNNNNNNSCGCGC-----CCCCGCCNNNNNNNSGCCGCCGCCGNNNNNNNSCGC 108690
Qy 4965 tggcccccaacccccacccacccacccacccacccacccacccacccacccacccaccc 5024
Db 108689 CG 108630
Qy 5025 acacggcgcgctgagaaacggagacacacacacacacacacacacacacacacacacacac 5084
Db 108629 CCNNNNNSAGCN 108570
Qy 5085 tgcacacacacgcccacccgcatggcccagcgagctgatatgctgagggcgctctcgc 5144
Db 108569 NSCCAGGCG 108510
Qy 5145 cccgagtcctgctggaactaaactagctgcgggtcccgagggacatcatcgacctgt 5204
Db 108509 GCCAGGCG 108450
Qy 5205 cccaagtccac 5264
Db 108449 CGAGCCAGCN 108391
Qy 5265 tggacgccttgcttacctccccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5324
Db 108390 NNNNSGGGAGCCAGGCG 108331
Qy 5325 cccactctccccagaggtccacacacacttgacaaaacacacacacacacacacacacacacac 5384
Db 108330 CCCCCGNNNNNNNSVAGCGAGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 108271
Qy 5385 agcgggagcagacgggagcagagcgggacgggagcgggagcgggagcgggagcgggagcggg 5444
Db 108270 CCAGCG 108211
Qy 5445 tcacgtccaccacgacggtggagcagcaccacacacacacacacacacacacacacacacac 5504
Db 108210 NNNNSCCCVAGCGAGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 108151
Qy 5505 gggcagcagcagcagcagcagcgggggtgggggagcagcagcagcagcagcagcagcagcagcag 5564
Db 108150 GCG 108091
Qy 5565 cccatgccaccac 5595
Db 108090 NNNNSGGGCGCVAGGCGAGCCAGGCGCGCG 108060

Search completed: September 8, 2001, 09:57:07
Job time: 1139 sec

1039	DB	GCTCTTGTACAGAGGGGAGTGGACCCCTGCTGAGCAAAAGGAGTGATTTCTCGATCACCAAG	1098
6790	QY	caacacacccagccgacgaccttcttcagcaagctgacggagagcaactccgccatgggt	6849
1099	DB	AAGTATAAGCTACTTGCCTTCATTCTTCCACCAAGCT--TGAAAGCACATCACCCATGGT	1155
6850	QY	caagtccaagaagaagagatacaacaagaagctgaacacccacacacccgggaatgacgtga	6909
1156	DB	TTAATCAAAAGAAACAGGAAATTTTTCGTAACTTGAATCTTCTGGTGGAGGTGACTCTGA	1215
6910	QY	atacaatatcagcagcgtggagcggagatcttcaatatcccgccatcacccggacaacag	6969
1216	DB	TATGCGAGTGTCTAGCCAGGACAGAGATCTTCAATCTGCCAGCATTTACCAACATCAG	1275
6970	QY	ctttatgacctatagaacggcggtgcagaaacatgccagcaccaacatggggctgga	7029
1276	DB	TGCAGTGAGCTCAAGAAGCCATTCTTTTGCTGATCCCGCCAG--TAACTTGGTCTAGA	1332
7030	QY	ggccataatagaaggcactcatgggtaaat---atgaccaagtgggaagagtcgccgc	7086
1333	DB	AGACATCATCAAGAAAGGCTCTCATGGGAAGTTTGTATGATAAAGTTGAAGATCATGGTGT	1392
7087	QY	gctcagcgccaatgctttaacctctgaatgccagtgccagctgcgcgtcgtatgcc	7146
1393	DB	TGTCATGTCCCCATCTGTGGGCATTATGCCGTGTAGTGCACCACTTCAGTGTGACGAG	1452
7147	QY	cataacccgtctgcagcagcagtgaccacacactcacctcgccaggtggcgcgaggaa	7206
1453	DB	CAGCGAGGCACGAGAGATGAAGGGAGCCATACCTCATCCAGGAGTATGCAACCAAA	1512
7207	QY	ggcccaaggtctctggcagaccagccgaaagccaagtcccccggcccgggcctggc	7266
1513	DB	GCTGATCAACAATCAACACAGCAGGAAGCTCTTAAATCTCTATTCTCTGGGCAAAAGCTATT	1572
7267	QY	alctgggacccgaccacctctgtctctcagtcagctcagagggagagctgaacccgcg	7326
1573	DB	AGGAAGTGAAGGCCTCTTCTGTCTCTGTGCATCAGAAGTGATTACCACAGGCA	1632
7327	QY	gagcgcgctcaccaaacccgctgtgggagagacaggccctcgctccgcagaggttcccacgcatt	7386
1633	DB	GACAC-----CAGGATGGGCATGGGAAGATCGGCCCTCTTCAACAGGTTCTACTCAGTT	1686
7387	QY	ccccataacccccctgatcgtggctgcaggcgggtgtcatggcttccccacccccacc	7446
1687	DB	CCCTTACAACCCCTGTACCATACGGATGCTCAGCAGTACACCACTTACACAGATCCGATG	1746
7447	QY	ggcgcttcccgcggggcagc---ggggccccctcgctggccccacacccctcgggacgaga	7503
1747	DB	CGCCCATCTGCCATCACCACAGAGCTCCACATCAACAGAACCGCATCTGGGAGAGGA	1806
7504	QY	ggccagccactgtctctgcagtagtacgacactctccgacagoga	7551
1807	DB	GCTGCCCGCTTCCTCAGCGCAGTATGACACACTGCTGTATGATGA	1854

RESULT

```

PCIT-US95-16311-9
; Sequence 9, Application PC/TUS9516311
; GENERAL INFORMATION:
; APPLICANT: Moore, David
; APPLICANT: Seol, Wongi
; APPLICANT: Choi, Hwang-Sik
; TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
; TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street, Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2001, 09:38:08 ; Search time 5930.11 Seconds
(without alignments)
13662.547 Million cell updates/sec

Title: US-09-522-753-4
Perfect score: 8571
Sequence: 1 catgtcgggtccacacagc.....caaaaaaaaaaaaaaaaaaa 8571

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1:	gb_est1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_est4:*
5:	gb_est5:*
6:	gb_est6:*
7:	gb_est7:*
8:	gb_est8:*
9:	gb_est9:*
10:	gb_est10:*
11:	gb_est11:*
12:	gb_est12:*
13:	gb_est13:*
14:	gb_est14:*
15:	gb_est15:*
16:	gb_est16:*
17:	gb_est17:*
18:	gb_est18:*
19:	gb_est19:*
20:	gb_est20:*
21:	gb_est21:*
22:	gb_est22:*
23:	gb_est23:*
24:	gb_est24:*
25:	gb_est25:*
26:	gb_est26:*
27:	gb_est27:*
28:	gb_est28:*
29:	gb_est29:*
30:	gb_est30:*
31:	gb_est31:*
32:	gb_est32:*
33:	em_estba:*
34:	em_estfun:*
35:	em_esthum1:*
36:	em_esthum2:*
37:	em_esthum3:*
38:	em_esthum4:*
39:	em_esthum5:*
40:	em_esthum6:*
41:	em_esthum7:*
42:	em_esthum8:*
43:	em_esthum9:*
44:	em_esthum10:*
45:	em_esthum11:*
46:	em_esthum12:*
47:	em_esthum13:*
48:	em_esthum14:*
49:	em_esthum15:*
50:	em_esthum16:*
51:	em_esthum17:*
52:	em_esthum18:*
53:	em_esthum19:*
54:	em_esthum20:*
55:	em_esthum21:*
56:	em_esthum22:*
57:	em_esthum23:*
58:	em_esthum24:*
59:	em_esthum25:*
60:	em_esthum26:*
61:	em_esthum27:*
62:	em_esthum28:*
63:	em_estin1:*
64:	em_estin2:*
65:	em_estin3:*
66:	em_estin4:*
67:	em_estin5:*
68:	em_estom1:*
69:	em_estom2:*
70:	em_estov1:*
71:	em_estov2:*
72:	em_estpl1:*
73:	em_estpl2:*
74:	em_estpl3:*
75:	em_estpl4:*
76:	em_estpl5:*
77:	em_estpl6:*
78:	em_estpl7:*
79:	em_estpl8:*
80:	em_estpl9:*
81:	em_estpl10:*
82:	em_estro1:*
83:	em_estro2:*
84:	em_estro3:*
85:	em_estro4:*
86:	em_estro5:*
87:	em_estro6:*
88:	em_estro7:*
89:	em_estro8:*
90:	em_estro9:*
91:	em_estro10:*
92:	em_estro11:*
93:	em_estro12:*
94:	em_estro13:*
95:	em_estro14:*
96:	em_estro15:*
97:	em_estro16:*
98:	em_estro17:*
99:	em_estro18:*
100:	em_estro19:*
101:	em_estro20:*
102:	gb_est25:*
103:	gb_est26:*
104:	gb_est27:*
105:	gb_est28:*
106:	gb_est29:*
107:	gb_est30:*
108:	gb_est31:*
109:	gb_est32:*
110:	gb_est41:*
111:	gb_est42:*
112:	gb_est43:*
113:	gb_est44:*
114:	gb_est45:*
115:	gb_est46:*
116:	gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	703	8.2	876	175	BG252161	BG252161 602365028
2	679.6	7.9	1087	175	BG252257	BG252257 602365136
3	674.2	7.9	787	140	BE793487	BE793487 601588814
4	668.4	7.8	804	139	BE728145	BE728145 601563413
5	661	7.7	700	141	BE900740	BE900740 601673910
6	656.8	7.7	1111	147	BF307042	BF307042 601898975
7	645	7.5	746	150	BF529782	BF529782 602044067
8	632	7.4	660	165	BE298158	BE298158 601118532
9	630.4	7.4	688	167	BE391091	BE391091 601286040
c 10	628.2	7.3	657	150	BF530324	BF530324 602071630
11	622.6	7.3	920	174	BG117169	BG117169 602346448
12	619.4	7.2	622	102	AI830862	AI830862 wj61b09.x
13	613.8	7.2	629	140	BE795997	BE795997 601591169
14	609.2	7.1	650	165	BE264455	BE264455 601191770
15	602	7.0	624	153	BG391632	BG391632 602417893
c 16	599.4	7.0	664	143	BF058781	BF058781 7k34c07.s
c 17	594.8	6.9	731	13	AA872713	AA872713 ob11b07.x
18	592	6.9	592	148	BF436351	BF436351 7p06e11.x
19	587	6.8	611	116	AW438580	AW438580 xs82e10.x
c 20	585.4	6.8	902	154	BG485444	BG485444 602505328
c 21	584.6	6.8	612	21	AI523558	AI523558 th08c09.x
22	584.4	6.8	922	136	BE542336	BE542336 601067079
23	580.4	6.8	610	138	BE675456	BE675456 7109d11.x
24	578.4	6.7	612	119	AW674918	AW674918 bd30f04.y
c 25	575.8	6.7	619	149	BF475570	BF475570 nac4e06.
c 26	571.4	6.7	1030	174	BG119261	BG119261 602349134
c 27	558.4	6.5	762	19	AI373205	AI373205 qz13h02.x
28	549.6	6.4	916	153	BG420527	BG420527 602452582
29	548	6.4	548	144	BF062632	BF062632 7h62h12.x
30	543	6.3	559	119	AW674586	AW674586 bd34f04.y
31	540.8	6.3	608	140	BE399873	BE399873 QV0-FN018
c 32	538	6.3	665	12	AA837468	AA837468 od20e09.s
33	533	6.2	665	141	BE903379	BE903379 601676011
c 34	531.2	6.2	777	21	AI499095	AI499095 t04f05.x
35	526.2	6.1	543	165	BE264249	BE264249 601191857
36	524.6	6.1	549	122	AW950948	AW950948 EST363018
c 37	524.4	6.1	623	102	AI812033	AI812033 tw46d03.x
c 38	522.4	6.1	537	24	AI761007	AI761007 wh97d05.x
c 39	520.8	6.1	538	102	AI831661	AI831661 wj50e06.x
c 40	513.8	6.0	610	150	BF527658	BF527658 602040636
41	511	6.0	511	172	BG025574	BG025574 602274985
c 42	510.8	6.0	551	7	AA421597	AA421597 zu28d02.s
c 43	508.6	5.9	558	122	AW956613	AW956613 EST368683
44	503	5.9	642	150	BF528779	BF528779 602041380
45	500.4	5.8	526	137	BE548650	BE548650 601074545

ALIGNMENTS

RESULT 1
 LOCUS BG252161 876 bp mRNA EST 13-FEB-2001
 DEFINITION 602365028F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4473271 5', mRNA sequence.
 ACCESSION BG252161
 VERSION BG252161.1 GI:12761977
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 876)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999).
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM10295 row: h column: 08
 High quality sequence stop: 711.
 Location/Qualifiers
 1. 876
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4473271"
 /clone_lib="NIH_MGC_90"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site: 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
 BASE COUNT 188 a 322 c 257 g 109 t
 ORIGIN

Query Match 8.2%; Score 703; DB 175; Length 876;
 Best Local Similarity 93.9%; Pred. No. 1e-121;
 Matches 831; Conservative 0; Mismatches 40; Indels 14; Gaps 9;
 QY 3272 actgcccgcgcgtctgcccgcgcacccacccatctccaccccccctccctccctcc 3331
 Db 1 ACTGCGCGCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
 QY 3332 ttgtgcaagcagcccgagctctctgagagggcaaatagtgccatctcccaaggaatgtcg 3391
 Db 61 TCTGCGCAAGCAGCCCGAGCGTCTCGAGAGGCAATAGGTGCCATCTCCCAAGGAATGCG 120
 QY 3392 gtccagctccagctccgctactcagagatgcagagcccgctgggctccctccaccatg 3451
 Db 121 GTCCAGCTCCAGCTCCCGTACTCAGAGCATGCAAGGCGCGCGCGCGCGCGCGCG 180
 QY 3452 gggctgcccctgcccagcagcccccaccccaaaagctggcacccttcagcgagtgagcagag 3511
 Db 181 GGGCTGCCCTGCCCATGGACCCCAAAAGCTGGCACCCCTTCAGCGGAGTGAAGCAGGAG 240
 QY 3512 cagctgtcccccacgggcccagctgggcccacccggagagcctgggggtgcccacagcccag 3571
 Db 241 CAGCTGTCCCCACGGGGCGCAGGCTGGGGCACCGGAGAGCGCTGGGGGTGCCACGCCAG 300
 QY 3572 gagcgctcgtgctgagagggacagctctgggctcagttccgggcccgaagcaccacaa 3631
 Db 301 GAGGCGTCCGTGCTGAGAGGGACAGCTCTGGGCTCAGTTCCGGGCGGAAGCATCACCAA 360
 QY 3632 ggcattcccagcacacggggtgcccctcgacagcgccatcacatcccggtgccatccacc 3691
 Db 361 GGCATTCCAGCACACGGGTGCGCTCGGACAGCGCCATCACATCCGGGCTCCATCAC 420
 QY 3692 cagcgacgcagctgagctcctgtacaagggcacacacacacacacacacacacacacac 3751
 Db 421 CACGGCAGCCAGCTGAGCTCTGTGTACAGGGGACCATCATCAGGATCATCGCGGAGGAC 480
 QY 3752 agcccagctcgtgagcggcgccgggagagacagccctgcccaaggccacacacacacac 3811
 Db 481 AGCCCGAGTCGGTCTGGACCGCGCGGGGAGGACAGCTGCCCAAGGCCACCTCATCTAC 540
 QY 3812 gaagcagaagggccacagctctgtcctatgaggtggcagctgtgtaccagctgtctcc 3871
 Db 541 GAAGGCAAGAGGGGCCAGCTCTGTGTCTTATGAGGTGGCATCTGTGACCCAGTGTCTCC 600
 QY 3872 aagggagcgcgcagagcagctcagagccccccatgagacggcgcccccagcgcacc 3931
 Db 3872 aagggagcgcgcagagcagctcagagccccccatgagacggcgcccccagcgcacc 3931

TITLE
JOURNAL
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM797 row: 1 column: 04
High quality sequence stop: 762.

FEATURES
source
1. 787
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3943059"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 187 a 248 c 231 g 121 t
ORIGIN

Query Match 7.9%; Score 674.2; DB 140; Length 787;
Best Local Similarity 95.5%; Pred. No. 2.6e-116;
Matches 748; Conservative 0; Mismatches 28; Indels 7; Gaps 5;

Qy 6645 tgggtggtggtgagcggattgaacctgtgtcccaaccgagggcattgacgagccag 6704
Db 5 TGGGTGTTGTTGAGCAGCGTATTGAACCTGTGTCCCAACCGAGGCGATGACGGAGCCA- 63

Qy 6705 ggcactcccgagtgctgtgtaccgctgtgtaccggtatggggagacagcagagccca 6764
Db 64 GGCACATCCCGGAGTGCTGTGTACCGCTGTGTACCGGGATGGGAGACAGACGGAGCCCA 123

Qy 6765 gcagatggtctcaagtctccagcgaacacacagccagcccgccctctt-cagcaag 6823
Db 124 GCAGGATGGGTCTCAAGTCTCCAGGCACACACAGCCAGCCGCCAGCCCTTCTTACAGCAAG 183

Qy 6824 ctgaccgagagcaactccgcatggtcaagtccaagaagcagagatcaacaagaagctg 6883
Db 184 CTGACCGAGAGCAACTCCGCCATGCTCAAGTCCAAGAGCAAGAGATCAACAAGAAGCTG 243

Qy 6884 aacacccacacccgaatgagcctgaatacataatcagccagcctggagcagagattctc 6943
Db 244 AACACCCACACCGGAATGAGCCTGAATACAAATATCAGCCAGCCTGGGACGGAGATCTTC 303

Qy 6944 aatgcccgcacacacggaacagcccttatgacctatagaccagcagcggtcagaaa 7003
Db 304 AATATGCCCGCCATACCGGAACAGGCCCTTATGACCTATAGAAGCCAGCGGTGAGGAA 363

Qy 7004 catgccagcacaacatgggctgagggccataatagaaggccactcatggtgataat 7063
Db 364 CATGCCAGCACCACATGGGCTGGAGGCCATAATTAGAAGGCACATCATGGTAAATAT 423

Qy 7064 gaccagtgggaagatgcccgcctcagcgccaatgcttttaacccctctgaatgccagt 7123
Db 424 GACCAGTGGGAAGATGCCCGCCCTCAGCGCCAATGCTTTAAACCCCTCTGAATGCCAGT 483

Qy 7124 gcaagcctgcccgtgtatgtcccatcaacccgtctgacgagcaggtgaccacacactc 7183
Db 484 GCCAGCCTGCCCTGTGTATGCCCATAAACCGCTGCTACGACGACGGAGTGACACACATC 543

Qy 7184 acctgccaggtgcccgggaagccaaagtctctgacagccagcagccgaaagcc 7243
Db 544 ACCTGCCAGTGGGGGGGAGGCCAAGCTCTGCGACAGCCAGCAGCGAAAGCC 603

Qy 7244 aagtcctccgcccggcctggtgcatctgggagccgcccacccctctgtctctcagtgcaac 7303
Db 604 AAGTCCCGGGCCGGCCTGGCATCTGGGAGCCGG-CACCCCTGTCTCTCTCAATGTCAC 662

Qy 7304 tcgagggagagactcaaccgcccgcgcgcctcaccacaccgctgtggagagacagcgg 7363
Db 663 TCGAGGAGAGACTCAACCCGC--GGACCCCTCAACCAACCCGGTGTGGGAGGACAGGCC 720

Qy 7364 tcgtccgaggttccacgcccattccctcacacccctgacatcgcgctcagcgcggt 7423
Db 721 TCGTCCGCAG--TTACGCCCATTCCTAAATCCCTGATCATCGGTGCGAGCGTGCAGGG 778

Qy 7424 gtc 7426
Db 779 TTC 781

RESULT 4
BE728145
LOCUS BE728145 804 bp mRNA EST 15-SEP-2000
DEFINITION 601563413F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3833162 5', mRNA sequence.
ACCESSION BE728145
VERSION BE728145.1 GI:10142137
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 804)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM511 row: 1 column: 03
High quality sequence stop: 716.
Location/Qualifiers
1. 804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3833162"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 200 a 236 c 262 g 106 t
ORIGIN

Query Match 7.8%; Score 668.4; DB 139; Length 804;
Best Local Similarity 92.6%; Pred. No. 3.2e-115;
Matches 746; Conservative 0; Mismatches 56; Indels 4; Gaps 4;

Qy 242 gagctccacctcgccagagtcacactcatacctgcccgagctggggaagtcagagatg 301
|||||

Db	599	CTGCTTACCGTTCAGAGCGGTGTGCAGCGCAGAGTGGTACCGCCATCACTCGCAGGA	658
Qy	8391	ccaaagggggcgggactgctcgtcacgcccgcgtgtgtc	8429
Db	659	CCAAGGGGGCGGAGCTGCTGGCTCACGCCCGCTGTGTC	697
RESULT	6		
BF307042			
LOCUS	60189875f1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4123770 5',	EST	21-NOV-2000
DEFINITION	mRNA sequence.		
ACCESSION	BF307042		
VERSION	BF307042.1	GI:11254150	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1111)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution Information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov Plate: L1CMI009 row: e column: 19 High quality sequence stop: 629.		
FEATURES	Location/Qualifiers		
source	1..1111		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4123770"		
	/clone_lib="NIH_MGC_17"		
	/tissue_type="rhaddomyosarcoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;		
	Site_2: XhoI; cDNA made by oligo-dT priming.		
	Directionally cloned into EcoRI/XhoI sites using the		
	following 5' adaptor: GGCACGAG(G). Size-selected >500bp		
	for average insert size 1.8kb. Library constructed by		
	Ling Hong in the laboratory of Gerald M. Rubin (University		
	of California, Berkeley) using ZAP-cDNA synthesis kit		
	(Stratagene) and Superscript II RT (Life Technologies)."		
BASE COUNT	246 a 321 c 344 g 200 t		
ORIGIN			
Query Match	7.7%; Score 656.8; DB 147; Length 1111;		
Best Local Similarity	89.1%; Pred. No. 4.8e-113;		
Matches	774; Conservative 3; Mismatches 77; Indels 15; Gaps 6;		
Qy	7590	sperlutcaagttccacgagcacaggaacgccttcgagggcggcgtgcgga 7649	
Db	1	GGCGGTGTCAAGTCCACGACGACACAGAACGGCCCTGCAGGACGAGGGGGTGC	60
Qy	7650	ctcccccaacgaaggagcccttaagctcgctccatccatctgcgtcc 7709	
Db	61	CTCCCCAACCAAGAAAGAGGCCCTGAGTCGGCTTCATCATCTGTCGGTCC 120	
Qy	7710	agagccggcatccttgcctgtctaaagccttaactaagactccgcggcgtggccct 7769	
Db	121	AGAGCGGCATCCTTGCTGTGCTAAAGCCTTAAGACTCCGCCCGGGCTGCCT 180	
Qy	7770	gtcagaccttactcagggggaagttaacctgggtgctcgggaaggaggggagggccgg 7829	
Db	181	GTGACAGACCTTACTCAGGGATGTTTACTCGTGTCTGGGAAGGAGGAGGGGCGG 240	


```

Db 301 CTGGCTCCCGCCGAATGCAATTTGGAAACCAAAAGTCTAAACTGAGCTCGACGCCCGCCGCGCC 360
Qy 7980 tccctcgcctcccatcccgcttagcgtctgagacagatggagcagcgagccctgtccagcc 8039
Db 361 TCCTCCGCGCTCCCATCCGCTTAGCGCTCTGGACAGATGGACGAGCGCCCTGTCCAGCC 420
Qy 8040 cccagtgcgtcttcctccggtcccccacagactgccccagcccaacagagatgctggaaacca 8099
Db 421 CCCAGTGGCGCTGTTCCGCTCCCGCCACAGACTGCCCCAGCAACAGAGATGGCTGGAACCA 480
Qy 8100 actcagggcaggtggcgagacaaaagggccaggtgcgcctggggggaacgagatctccg 8159
Db 481 AGTCAGGCGCAGTGGCGGACAAAAGGGCCAGGTGCGGCGCTGGGGGGAACGAGATGCTCG 540
Qy 8160 aggaactggaactgtttttttcacacatcgtgcccgcagcggtgggaagaaagcgagatgt 8219
Db 541 AGGACTGGACTGTTTTTTCACACATCGTTGCGCAGCGGTGGGAGCGACGCCGCGATGT 600
Qy 8220 a-aatgatgtgtgtttacaggggtatatattttgtataccttcaatgaatgaattcagatg 8278
Db 601 ACCCTGATGTGTGTTACAGGCTATATCTTTGATACCTTCATGAATTAATTCAGATG 660

RESULT 9
BE391091 688 bp mRNA EST 21-JUL-2000
LOCUS 601286040F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3607929 5',
DEFINITION mRNA sequence.
ACCESSION BE391091
VERSION BE391091.1 GI:9336456
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 688)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM259 row: p column: 10
High quality sequence start: 8
High quality sequence stop: 684.
FEATURES
Location/Qualifiers
Source 1..688
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3607929"
/clone.lib="NIH_MGC_44"
/tissue_type="endometrium, adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

168 a

213 c

208 g

99 t

Query Match

Best Local Similarity 7.4%; Score 630.4; DB 167; Length 688;

Matches 675; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

```

Qy 245 ctccacctggccagagtcacactcactgcccagctgcccgaagtcagagatggag 304
Db 6 CTCCACCTGGCGCCAGAGTCCCACTCATACCTGCCGAGCTGGGGAAGTCAGAGATGGAG 65
Qy 305 ttcaatgaagaagaagcgcctcggtgctgtagctgctgacccccctgctggaacgtca 364
Db 66 TTCAATTGAAGAACAAGCGCGCTCGGCTAGAGCTGCTGCTGACCCCTGCTGGACCGGTCA 125
Qy 365 cccctgctggccagggccagcctcggtggtgagacctcacaagagacgtgacctg 424
Db 126 CCCCCTGCTGGCCACCGGCGAGCCTGCGGATCTGAAGACCTCACCAGGACCGTAGCCCTG 185
Qy 425 acgggcaagctggaacccggtgtctccccccagccccccgcacactgacctgagctggag 484
Db 186 ACGGGCAAGCTGGACCGGTGTCTCCCCCCCAGCCCGCCGACACTGACCTGAGCTGGAG 245
Qy 485 ctggtgcgcgcacggtgtcccaaggagagctgtatccagaacatggaacgcgtgaccga 544
Db 246 CTGGTCCGCCACGGCTGTCCAAGGAGGAGCTGATCCAGAACATGACCGCTGGACCGCA 305
Qy 545 gagatcaccatgtagagcagcagatctctaagctgaagaagaagcagcaacagctggag 604
Db 306 GAGATCACCATGTTAGAGCAGCAGATCTCTAAGCTGAAGAAGACAGCAACAGCTGGAG 365
Qy 605 gagggagctgccaagccgcccgcgctgagaagccgctgtcaccgcccctcagctcagtcg 664
Db 366 GAGGAGGCTGCCAAGCCCGCGAGCTGTGAGAGCCCGTGTACCCGCCGCCATCGAGTGC 425
Qy 665 aagcaccgcagcctggtgcagatcatctacgacgagaaacccggaagaagctgaaactgca 724
Db 426 AAGCACCGCAGCCTGGTGAGATCATCTACGACGAGAACCGGAAGAGCTGAAGCTGCA 485
Qy 725 catcggattctggaagcctggggcccgagctggagctgcccgtgtataaacagccctcc 784
Db 486 CATCGGATTTCTGGAAGGCTGGGGGCCCGCAGGTGGAGCTGCCCTGTGTACAAACAGCCCTCC 545
Qy 785 gacacccgcagctatcatgagacatcaataaaacccagcgctggaagaagctaatc 844
Db 546 GACACCCGGCAGTATCATGAGAACATC-ANATPAAACAGCGCATGCGGAAGAAGTAATC 604
Qy 845 ttgtacttcaagagaggaatcacgcctcggaacaaatggaagcagaagtctgcagcgc 904
Db 605 TTGTACTTCAAGAGGAGG-ATCACGCTCGGAAC-ATGGGAGCAGAG-TCTGCCAGCGC 661
Qy 905 tatgaccagctcagagcgc 924
Db 662 TATGACCAGCTCATGGAGGC 681
```

RESULT 10

BE391091/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BF530324 657 bp mRNA EST 11-DEC-2000

602071630F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4214551

5', mRNA sequence.

BF530324

BF530324.1 GI:11617698

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 657)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: David N. Louis, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9788 row: 1 column: 08

High quality sequence stop: 657.

FEATURES

source

1..657

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:4214551"

/clone_lib="NCI CGAP Brn67"

/tissue_type="anaplastic oligodendroglioma with lp/19q

loss"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 2.3 kb. Constructed by Life

Technologies. Note: this is a NCI CGAP Library."

85 a 195 c 263 g 114 t

BASE COUNT

ORIGIN

Query Match 7.3%; Score 628.2; DB 150; Length 657;

Best Local Similarity 99.5%; Pred. No. 1.1e-107;

Matches 630; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4340 aaggagggtccatcacgcaggggaccccgctcaagtacgacccggcgctccaccact 4399

Db 657 AAGGAGGGTCCATCAGCAGGGACCCCGCTCAAGTACGACACCGCGCGTCCACCACT 598

QY 4400 ggtccaaaaagcagcagctagctccctcatcgagcagccggcgagctccacc 4459

Db 597 GGTCTCAAAAACACGACATAGCTCTCTCTCATCGCAGCCCGCGCGGACGTTCCACACC 538

QY 4460 gtgcacccgtgattgtatgctgcagcagccggcgacgtgaagcgtgctgctacagag 4519

Db 537 GTGCACCCGCTGGATGTATGTCGCGGACCGCGGCGGACGTCGCTGCTTACGAGAG 478

QY 4520 agcctgaagagccggcagggaccccgacagcagctcggggggtccattgctgcgctgccc 4579

Db 477 AGCCTGAAGAGCGCGCAGGAGCGCGCAGCAGCTCGGGGGGTCCATTGCGCGCGCGGCC 418

QY 4580 ccggtcattgtcctgagctggttaagccgagcggcgagagcccccctgacctatgagaccac 4639

Db 417 CCGGTCTATTGTGCTGAGCTGGGCAAGCGCGGACAGCCCGCTAACCTATGAGGACCAC 358

QY 4640 ggggcacccctttgcccggccacctccacagaggttcgcccgtgacctgagcgggagccacg 4699

Db 357 GGGGACCCCTTTGCCGGCCACCTCCACGAGGTTGCGCGGTGACCCACGCGGGAGGCCACG 298

QY 4700 ccgctcgtcagaggagcgcctttgctccagaggcatcccgagaccgaaagctgagc 4759

Db 297 CCGCGCTCGAGGAGGCGCAGCCTTTGCTCAGCAAGGATCCAGGACCGAAAGCTGAGC 238

QY 4760 tcgagcctctgagatgcgcaagtccccgacacgacacccgtgcccagaccaccaccacac 4819

Db 237 TCGAGCGCTCGTGAGATCGCCAAAGTCCCGCACACACCGCTGCCGAGCACCACCCACAC 178

QY 4820 cccattcgcctatgagcacctgtctggggcgctgagtgccgtgacctgctatcgagc 4879

Db 177 CCGATCTGCCCTATGAGGACCTGCTTGGGGCGGTGAGTGGGGGTGACCTGTATCGCAGC 118

QY 4880 cacatccccctggccttgacccccaccctccatcccccgcggtccctcttgagcagacc 4939

Db 117 CACATCCCGCTGGCCTTCGACCCACCTCCATACCCCGCGGATCCCTCTGAGCAGCC 58

QY 4940 gctgctactactgccccgacacactggcccc 4972

Db 57 GCTGCTACTACTCTGCCCGACACCTGGCCCCC 25

RESULT 11

BG117169

LOCUS

BG117169

920 bp

mRNA

EST

30-JAN-2001

DEFINITION

602346448F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4441595 5',

mRNA sequence.

ACCESSION BG117169

VERSION BG117169.1 GI:12610675

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC http://mgc.nci.nih.gov/.

1 (bases 1 to 920)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cagbbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM10212 row: p column: 12

High quality sequence stop: 741.

Location/Qualifiers

1..920

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4441595"

/clone_lib="NIH_MGC_90"

/lab_host="DH10B (phage-resistant)"

/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 1.7 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC Library."

204 a 277 c 266 g 173 t

BASE COUNT

ORIGIN

Query Match 7.3%; Score 622.6; DB 174; Length 920;

Best Local Similarity 92.1%; Pred. No. 1.2e-106;

Matches 778; Conservative 0; Mismatches 54; Indels 13; Gaps 11;

QY 7640 cagctgcgcagctcccccaacgaagagagccctgagctgcgcctccatccat 7699

Db 1 CGGCTGCCGACTCCCGCAACCAAGGAAGGAGCCCTGAGTCCGCTGCCCTCCATCAT 60

QY 7700 ctgtcgtccagagcggcagctcctgctgtctaaagccttaactaaagactccgcccgcg 7759

Db 61 CTGTCCGTCCAGAGCGGCATCTTGCCTGTCTAAGGCTTAAGTAAAGCTCCGCCCG 120

QY 7760 ggcctggccctgtgagacaccttactactgagggatgtttacctggtgctcggaagagggg 7819

Db 121 GCGTGGCCCTGTGCAGACCTTACTCAGGGATGTTTACCTGTGCTCGGGAAGGAGGGG 180

QY 7820 aagggccgggagggggcagcggcagcgtgtggcagccacacacagcggcggcggcggg 7879

Db 181 AAGGGCCGGGAGGGGGCAGCGGAGGCGTGTGCGAGCACACAGCGCGGCCAGG--CG 238

QY 7880 ccaggagcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7939

Db 239 GCCAGGACCAAGACGAGGATGACACGACCTCCACGCACTGCTCCCGCAATGAT 298

QY 7940 tggacacaaagtctaaactgagctgagcagcccgcccgccctcccgctcccgctcccgctcccg 7999

Db 299 TGAACCAAAAGTCTAAAGTACGCTGAGCTCGCAG-CCCCGGCGCCCTCCCTCCG-CTCCCATCCG 356

QY 8000 cttagcgtctggacagatgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 8059

Db 357 CTTAGCGCTCTGGACAGATGACGAGCCCTGTCTCAGCCCCCAGTGGCTGCTTCCGGT 416

```
QY 8060 cccacagactccccagcaacagagattgctgggaacacaaagtcaggcaggtggcgga 8119
Db 417 CCCACAGACTGCCCCAGCAGCAGAGATTGCTGGAACCAACCAAGTCAGGCCAGTGGCGGA 476
QY 8120 caaaagggccaggtgctgctgggggaacgagatgctccgaggaactgactg-ttttttt 8178
Db 477 CAAAAGGGCCAGGTGCGGCTGGGGGGAACGAGATGCTCCGAGGACTGGACTGTTTTTTT 536
QY 8179 cacacatgctgcccagcagcgggtggaaagaaagcagatgt-aaatgatgtttggt-tt 8236
Db 537 CACACATGTTGCCAGCGGTGGGAAGAAAGGCAGATGTAATAATGATGTGTTGTCAT 596
QY 8237 acagggtatttttgataccttcaatgaattaattcag--atgttttacgcaagaa-g 8293
Db 597 CCAGGTATATTTTGATACCTTCAATGAATTAATTCAGGATGTTTTTTACGCAAGGAAG 656
QY 8294 gacttaccagattactgctgctgtgttttgatctctgttaccgttcaag-aggcgt 8352
Db 657 GACTTACCAGTATTACTGGTGCTGTGCTTTGATCTCATGTTCATGTTTCAAGAAGGCGT 716
QY 8353 gtgcagggccagactcgtgta-ccccatcactcgcagaccagggcgggagctgctc 8411
Db 717 TTGCAGGGCCGACAGTCGTGACCCCAATCAATCGCAGACCAAGGGG-GGGGACTGGTG 775
QY 8412 gtacgcccgcgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 8471
Db 776 GCTCAGGCCCGGTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 835
QY 8472 ttctg 8476
Db 836 ATCTG 840
```

RESULT 12

```
A1830862
LOCUS wj61b09.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2407289 3'
DEFINITION similar to TR:000613 000613 T3 RECEPTOR-ASSOCIATING COFACTOR-1. [1]
; mRNA sequence.
ACCESSION A1830862
VERSION A1830862.1 GI:5451533
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 622)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the i.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 2006 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 465.
FEATURES
source
1..622
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2407289"
/clone_lib="NCI_CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
```

```
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 138 a 228 c 171 g 85 t
ORIGIN
Query Match 7.2%; Score 619.4; DB 102; Length 622;
Best Local Similarity 99.8%; Pred. No. 4.7e-106;
Matches 620; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6349 cagccgcgtctccagaccgccccaggggtcaaaagtcaccagcggtgtgtcacccctgac 6408
Db 1 CAGCCCGCTGCTCCAGACGCGCCCGGGGTCAAAAGGTCAACAGCGGTGCTCACCCTGGC 60
QY 6409 ccagcacatcagtgaggtcatcacagagactacaccccgccacccacacagcagctcag 6468
Db 61 CCAGCACATCAGTAGGTGTCATCACAGGACTTACACCGGCACCCACAGCAGCTCAG 120
QY 6469 cgcacccctgcgcgcgcgcctctactctcctcctcctcctcctcctcctcctcctcct 6528
Db 121 CGCACCTCTGCGCGCGCCCTCTACTCTCTCTGCGGCAGCTGCGCCGCTCTGGACCT 180
QY 6529 ccgcgcgcacccagtgacctctacctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6588
Db 181 CGCGCGCCACCCAGTAGTACCTTACTCTCCGCCCGCGGACCATGTTGCCCGCCGCTGG 240
QY 6589 ctccccccacagcggaaggggcaagaggtctccagagccaaacaaagacgtcgtgtttg 6648
Db 241 CTCCCCCACAGCAGGAAGGGGCAAGAGGTCTCCAGAGCCAAACAGACGCTCGGTCTTGG 300
QY 6649 tgggtgagagcaggtattgaacctgtctccaccgagggcagcagcagcagcagcagcag 6708
Db 301 TGGTGGTGAGGACGGTATTGAACCTGTGTCCCCACCGGAGGCATGACGGAGCCAGGGCA 360
QY 6709 ctcccgagtgctgttaccgcgtgctgtaccggtggtggaacagcagcagcagcagcagc 6768
Db 361 CTCCTCGGAGTGTGTGTACCCGCTGCTGTGTACCGGATGGGGAACAGACAGGCCAGC 420
QY 6769 gatgggtccaaagtctccaggcaacaccagcagcagcagcagcagcagcagcagcagc 6828
Db 421 GATGGGCTCCAAAGTCTCCAGGCAACACCAGCCAGCGCCGAGCCTTCTTCAGCAAGCTG 480
QY 6829 cgaagcaactccgcctcagtgctcaagtcgaagcaagcaagcaagcaagcaagcaagca 6888
Db 481 CGAGAGCAACTCCGCCATGGTCAAGTCCAAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 540
QY 6889 ccacaacgcgaatgagcctgaatacaataatcagcagcagcagcagcagcagcagcagcag 6948
Db 541 CCACACCCGGAATGAGCTGTAATAATATCAGCCAGCCTGGGACGAGAGATCTTCAATAT 600
QY 6949 gcccgccatccaccggaacagg 6969
Db 601 GCCCGCCATCACCAGGACAGG 621
RESULT 13
BE795997
LOCUS BE795997 629 bp mRNA EST 20-SEP-2000
DEFINITION 601591169F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945305 5',
mRNA sequence.
ACCESSION BE795997
VERSION BE795997.1 GI:10217195
KEYWORDS EST.
SOURCE human.
```

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 629)
 REFERENCE NTH-MGC http://mgc.nci.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DTP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 plate: LLCM803 row: i column: 18
 High quality sequence stop: 629.
 Location/Qualifiers
 1..629
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3945305"
 /clone_lib="NIH_MGC_7"
 /tissue_type="small cell carcinoma"
 /cell_line="MGC3"
 /lab_host="DH10B (phage-resistant)"
 /note="Organelle: lung; Vector: pORF7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGACAGCAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 130 a 181 c 171 g 147 t
 ORIGIN
 Query Match 7.2%; Score 613.8; DB 140; Length 629;
 Best Local Similarity 99.5%; Pred. No. 5.3e-105;
 Matches 626; Conservative 0; Mismatches 2; Indels 1; Gaps
 QY 7922 gctctccccgaatgcatttggaaaccaaagtctaaactgagctgcagccccgcgcctc 7981
 Db 1 GCCTCCCCGAATGCATTGGACCAAAAGCTTAACATGAGCTCGACGCCCGCGCCCTC 60
 QY 7982 cctccgcctccccatcccgcttagctctgacagatggacgagccctgtccagcccc 8041
 Db 61 CCTCCGCTCCCATCCCGCTTAGCGCTCTGGACAGATGGACGAGGCCCTGTCTCAGCCCC 120
 QY 8042 cagtgcgctggttccggtccccacagactgccccagccaacagagattgctggaaacccaag 8101
 Db 121 CAGTGCCTCTGTTCCGGTCCCAACAGACTGCCCAAGCAACAGAGATTGCTGGAACCAACAG 180
 QY 8102 tcaggccaggtgggcggacaaaaggccaggtgcgccttgggggaacagatgctccag 8161
 Db 181 TCAGGCCAGTGGCGCGGCAAAAGGCCAGGTGCGGCTTGGGGGGNACGGATGCTCCGAG 240
 QY 8162 gactgactgtttttttcacacatcgttgcgcagcgtgggaagaaaggcagatgttaa 8221
 Db 241 GACTGGACTCTTTTTTTCACATCTGTTGCCGAGCGGTGGGAAGAAAGGCAGATGTAA 300
 QY 8222 atgatgttggtttacaggggtatatttttgataacctcaatgaattaatccagatgttt 8281
 Db 301 ATGATGTGTTGGTTTACAGGGTATATTTTGTGATACCTTCAATGAATTAATTCAGATGTT 360
 QY 8282 tacqcaaggaaggactaccagattactgcctgtgcttcttgcctctctctaccgt 8341
 Db 361 TAGCAGGAGGAGGACTATCCCAGTATTAATGCTGTGCTCTTTGATCTCTGCTTACCGT 420
 QY 8342 tcaagaggcgtgtgcaggccagtcgdtgacctcatcctcgacgaccaaaggggcg 8401
 Db 421 TCAAGAGGGGTGTGCAGGCGGACAGTCGGTGTACCCCATCACTCGCAGGACCAAGGGGCG 480

Db 180 CCACGACCTCCACGCCACTGCCTCCCGAATGCAATTTGGAACCAAGTCTAAACTGAG 239
QY 7962 ctgcagccccgcgcctccctccgcctcccatccgctttagcgtctggacagatga 8021
Db 240 CTGCGAGCCCCCGCGCCCTCCCTCCGCCCTCCCATCCCGCTTAGCGCTCTGGACAGATGA 299
QY 8022 cgcagggccctgtccagccccccagtcgctgttcggtgtcccccacagactgccccagccaa 8081
Db 300 CGCAGGCCCTGTCCAGCCCCCAGTGGCTCGTTCGGTCCCCACAGACTGCCCCAGCCAA 359
QY 8082 ctagattgtgaaacaaagtcaggccaggtg99gcgacaaaaggccaggtgcggcctg 8141
Db 360 CGAGATGTGTGAAACCAAGTCAGGCCAGGTGGCGGACAAAAGGCCCAAGTGCGGCCCTG 419
QY 8142 g99ggaaagatgtctccagggactgactgtttttttcacacatggttccgcagcggtg 8201
Db 420 GGGGAACGGATGCTCCGAGGACTGGACTGTTTTTTTTCACACATCGTTCGCCGCGGGTG 479
QY 8202 g9aagaaaggcagatgtaataga-tgtgtgtgtttacaggggtatatattttgtataccttc 8260
Db 480 GGAAGGAAAGGCAGATGTAATGACTGTGTCTCGTGTACAGGGTATATTTTTGTATACCTTC 539
QY 8261 aatgaattaatcagatgttttacgcaagaaggacttaccagttactactcgtctgtg 8320
Db 540 AATGAATTAATTCAGATGTTTTACGCAAGGAAGGACTTACCAGTATTTACTGCTGTGTG 599
QY 8321 ctttgatctctgtaccgttcaagagcggtgtgcagggccagacgctcg 8370
Db 600 CTTTGATCTCTGCTTACCGTTCAAGAGCGGTGTGCAGGCCCGACAGTCGG 649

RESULT 15
BG391632
LOCUS 602417893F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4537438 5',
DEFINITION mRNA sequence.
ACCESSION BG391632
VERSION BG391632.1 GI:13285080
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLNL0462 row: 1 column: 23
High quality sequence stop: 623.

FEATURES
source

1..624
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4537438"
/clone.lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
136 a 222 c 190 g 76 t

BASE COUNT
ORIGIN

Query Match 7.0%; Score 602; DB 153; Length 624;
Best Local Similarity 99.7%; Pred. No. 8.7e-103;
Matches 624; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 3660 acagcgccatcacataccggtccatccacacgagcagccagctgacgtcctgtaca 3719
Db 1 ACAGGCCCATACATACCGCGGCTCCATCACCCACGCGCAGCTGACGTCTCTGTACA 60
QY 3720 agggcaccatcacccaggatcatcgcgagagcagcccgagtcgttggaccggtccgg 3779
Db 61 AGGGCACCATCACAGGATCATCGCGAGGACAGACCCCGAGTCGTGTGGACCGCGCGG 120
QY 3780 aggaagcgtctcccaagggccagctcatcagaagcagaagggccagctcttctct 3839
Db 121 AGGACAGCTGTCCCAAGGGCCACGTCATCTACGAAGCAAGAGGCCACGCTTCTTCT 180
QY 3840 atgagggtggcatgtctgtgaccagtgctccaaaggagcagcagcagcagctcaggac 3899
Db 181 ATGAGGGTGGCATGTCTGTGACCCAGTGTCTCAAGGAGGACGCGCAGAGCTCAGGAC 240
QY 3900 cccccatagagc 3959
Db 241 CCCCCATGAGACG 300
QY 3960 gagccatctccacccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4019
Db 301 GAGCCATCTCTCAGCCAGCATCGAAGGTCTCATGGGCGCTGCCATCCCGCGGAGCGAC 360
QY 4020 acagccccccacacctcaagagcagcagcagcagcagcagcagcagcagcagcagcagc 4079
Db 361 ACAGCCCCCACCACCTCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 420
QY 4080 ctgcgtctcactgagggcacagggaggaactactcgtcggaggcagcagcagcagcagc 4139
Db 421 CTCGCTCTTACGTGGAGGCACAGGAGGACTACTCTGCTCGGAGGCGCAAGCTCTTAAGC 480
QY 4140 gggagggaagcgcctccgc 4199
Db 481 GGGAGGGCAGCGCTCCGCCCGCCCGCCCTCAGGGACCTGACCGAGGCGCTACAGACGC 540
QY 4200 aggcctggggcccttgaaagtgaagccggcccatgagggcctgggtggccacggtgaag 4259
Db 541 AGGCCCTGGCGCCCTGAAAGCTGAAGCCGG-CCATGAGGCGCTGTGTGGCCACGCGTGAAG 599
QY 4260 agcgggcgctccatccatgatc 4285
Db 600 AGGCGGGCGG-TCCATCCATGAGATC 624

Search completed: September 8, 2001, 14:32:15
Job time: 17647 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2001, 09:38:08 ; Search time 10979.9 Seconds
(without alignments)
12074.241 Million cell updates/sec

Title: US-09-522-753-4
Perfect score: 8571
Sequence: 1 catgtgggtccacacagc.....caaaaaaaaaaaaaaaaaa 8571

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_vil2:*
60: gb_vil3:*
61: gb_vil4:*
62: gb_vil5:*
63: gb_vil6:*
64: gb_vil7:*
65: gb_vil8:*
66: gb_vil9:*
67: gb_vil10:*
68: gb_vil11:*
69: gb_vil12:*
70: gb_vil13:*
71: gb_vil14:*
72: gb_vil15:*
73: gb_vil16:*
74: gb_vil17:*
75: gb_vil18:*
76: gb_vil19:*
77: gb_vil20:*
78: gb_vil21:*
79: gb_vil22:*
80: gb_vil23:*
81: gb_vil24:*
82: gb_vil25:*
83: gb_vil26:*
84: gb_vil27:*
85: gb_vil28:*
86: gb_vil29:*
87: gb_vil30:*
88: gb_vil31:*
89: gb_vil32:*
90: gb_vil33:*
91: gb_vil34:*
92: gb_vil35:*
93: gb_vil36:*
94: gb_vil37:*
95: gb_vil38:*
96: gb_vil39:*
97: gb_vil40:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	8541	99.6	8561	89	AF113003
2	8324.6	97.1	8686	89	AF125672
3	5455	63.6	5989	97	HSU37146
4	5037.2	58.8	8544	94	AF125671
5	4981.6	58.1	8388	94	AF113001
6	4378.2	51.1	7465	94	AF113002
7	2759.8	32.2	2842	91	BC004326
8	2618.6	30.6	2930	97	S83390

[illegible]

Db 2341 AGGCCACCCACCCACACGAGGAGACATCCCGGGGCCCCCATTTGAGCCACCCCGGCCCTC 2400
Qy 2401 tgaagccacggagcccttaagcccccacccagcaccccccacccctctgacacccctcc 2460
Db 2401 TGAAGCCACCGAGCCCTACGCCCCACACAGCACCCCATCGCCCTCTGACCTCTCTCC 2460
Qy 2461 tttgttccccaaaggaggagagagagagagagagagagagagagagagagagagag 2520
Db 2461 TGTGTTCCCAAGAGGAG 2520
Qy 2521 ggag 2580
Db 2521 GGAG 2580
Qy 2581 gcccctcaagagcagtgatgacagagagagagagagagagagagagagagagagag 2640
Db 2581 GCCCTCAAG 2640
Qy 2641 ggaagccgctgagggccacggccgagggggcgctcaaggcagagagagagagagag 2700
Db 2641 GGAGGCCGTGAGGCCACGGCCGAGGGGGCGCTCAAGCGAGAGAGAGAGAGAGAGAG 2700
Qy 2701 cggcagggccaccactgcaagagagctggggcgccccccagagagagagagagag 2760
Db 2701 CGGCAGGGGCCACCACTGCCAAGAGCTCGGGGGCGCCCCCAGAGAGAGAGAGAG 2760
Qy 2761 ctgcagtcagacagaggttgatgagggcgagggcgagagagagagagagagagag 2820
Db 2761 CTGCAAGTCAGACGAGGTGATGAGGCCGAGGGCGGCGACAGAGAACCGGCTGTG 2820
Qy 2821 aagggccagcctctcaccgccagctggcgacccccggggccaatgctcacccccagag 2880
Db 2821 AAGGCCAGCCTCTCTACCCCGACTGGCGAGCCCGCGGGCCCAATGCTCACCCGAG 2880
Qy 2881 actgagactgaagcagctgaagcagcagcagcagcagcagcagcagcagcagcag 2940
Db 2881 ACTGACCTGAAGCAGCTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2940
Qy 2941 agtccatgagccccccgggagagagagagagagagagagagagagagagagagag 3000
Db 2941 AGTCATGAGCCCCCGGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
Qy 3001 gccaccgcaaaccttcagcgcggagagagagagagagagagagagagagagagagag 3060
Db 3001 GCCACCGCAAAACCTGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
Qy 3061 gggcaagagcag 3120
Db 3061 GGGCAAGAGCAG 3120
Qy 3121 gaagctgctgggagaccccttctgtgacttcgggacttcgggacttcgggacttcgg 3180
Db 3121 GAAGCTGCTGGGAGACCCCTTCTGTGACTTCCGGCTGCGGCTGCGGCTGCGGCTG 3180
Qy 3181 tgagtgatcaaggctcccccgcagtcgccccggaocccctcagctctctactcagctcc 3240
Db 3181 TGAGTGATCAAGGCTCCCGCATGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Qy 3241 tggtaaccactgccccctgggctccatgacaactgcgcggccccctctgcgcggccac 3300
Db 3241 TGGTCAACCACTGGCCCTTGGGCTCCCATGACACTGCCCGGCCCTCTCTGCGCGCC 3300
Qy 3301 caccatctccaaacccgctccctctcactctctgcaagcaccacccagcgtccctcgag 3360
Db 3301 CACCATCTCCAAACCGCTCCCTCATCTCTCTGCAAGCAGCCAGCGTCTCTCGAGAG 3360
Qy 3361 gcaaataggtgccatctccccaaaggaatgctggtccagctccacgtcccgctactcag 3420
Db 3361 GCAATAGGTGCCATCTCCCAAGGAATTCGGTCCAGTCCAGTCCAGTCCCGTACTCAG 3420
Qy 3421 tgcgaagccccgagtgccctgacacatggagctgcccctgcccagtcgaccccaaaa 3480
Db 3421 TGCCAAGCCCCGGTGGGCCCTGTGACCATGAGGGGTGGCCCTTGCCCATGAGACCC 3480

Qy 3481 gctggcaccccttcagcggagtgaaagcagagagcagctgtcccccacggggccagctgg 3540
Db 3481 GCTGGCACCCCTTCAGCGGAGTGAAGCAGAGAGAGCTGTCCCCACGGGGCCAGGCTGG 3540
Qy 3541 accggagagcctgggggtgcccacagcccagggcgctcgtgtgagagggagacgtct 3600
Db 3541 ACCGGAGAGCCTGGGGTGGCCACAGCCACAGAGGCGTCCGTGTGAGAGGGACAGCT 3600
Qy 3601 gggctcagttccggggcgaagacatcaccaaaagggaattcccagcacacgggtgcctcg 3660
Db 3601 GGGCTCAGTTCCGGGCGGAAGCATCACAAAGGCAATTCACAGCACACGGGTGCCCTCG 3660
Qy 3661 cagcgccatcatatcccgcgctccatcaccacggcagcgcagctgagctgctgtacaa 3720
Db 3661 CAGCGCCATCATATACCGCGGCTCCATCACCCACGGCACGCCAGCTGACGTCTGTAC 3720
Qy 3721 gggcaccatcaccaggatcatcgcgagagacacccagctgcttgagacggcgccggg 3780
Db 3721 GGGCACCATCACAGGATCATCGCGAGGACAGCCGAGTCGCTTGGACCGCGCGGGGA 3780
Qy 3781 ggaagcctgcccgaaggcgcagtcactatcagaaggcaagaaggccacgtctgtccta 3840
Db 3781 GGACAGCCTGCCCAAGGGCCACGTCATCTACAAAGGCAAGAGGGCCACGCTTGTCT 3840
Qy 3841 tgagggtggcagtgctgtgacccagtgctccaaggagagcgcagaaagcagctcagga 3900
Db 3841 TGAGGGTGGCATGTCTGTGACCCAGTGCTCCAAAGGAGACGGCAGAGACAGCTCAG 3900
Qy 3901 ccccatgagagcggcgcccccaagcgcaactatgacatgagagggcgcggtggggcag 3960
Db 3901 CCCCATGAGAGCGCGCGCCCAAGCGCACCTATGACATGATGAGGGCGCGTGGGCGAG 3960
Qy 3961 agccatctctcagccagcagtcgaaggtctcatgggcgtgccaatcccgcggggagaca 4020
Db 3961 AGCCATCTCTCAGCCAGCATCGAAGGTCTCATGGCGCGTCCCATCCGCGCGAGCGCA 4020
Qy 4021 cagccccacacacctcaaaagcagcaccacatccgcgggtccatcacacaaggatccc 4080
Db 4021 CAGCCCCACCCACTCAAAAGAGAGAGACACCATCCGCGGGTCCATCACAAAGGGATCC 4080
Qy 4081 tcggtctctacgtggagcacagggagactacgtcgctgggagggcgaagctcctaaagcg 4140
Db 4081 TCGGTCTCTACGTGGAGGCACAGGAGGACTACTGCTGCGGAGGCCAAGCTCCTAAAGCG 4140
Qy 4141 ggaaggcagcctccgcccccaacgcgcctcagggacctcagggagggcgtacaaagcgca 4200
Db 4141 GGAGGGCACGCGCTCCGCCCCCACCGCCCTCACGGGACCTTGACCAGGCGCTACAAGAGCGCA 4200
Qy 4201 ggcctggggccctgaagctgaagccggcccagggcctggtggccacgggtgaagga 4260
Db 4201 GGCCCTGGGCCCCCTGAAGCTGAAGCGGGCCCATGAGGGCTGTTGGCCAGCGTGAAGGA 4260
Qy 4261 gggggcgcgctccatccatgagatcccgcgagagctgcggcacacgcccgcagctgccc 4320
Db 4261 GGGGGCGCGCTCCATCCATGATGATCCCGCGGAGAGGTGCGGCACACGCCCCAGCTGCC 4320
Qy 4321 cctggcccccgcgcgctcaaggagggtccatcagcagggggaaccccgctcaagtagca 4380
Db 4321 CCTGGCCCCGCGCGCTCAAGGAGGGCTCCATCAGCAGGGCACCCCGCTCAAGTAGCA 4380
Qy 4381 caccggcgcttccaccactggctccaaaagcacgacgtacgtccctcatcggcagccc 4440
Db 4381 CACCGCGCGTCCACCACTGGCTCCAAAAGCAGACGAGTACGCTCCCTCATCGCGACGCC 4440
Qy 4441 cggcgagcgttccccacccgtgcacccgtggaatgtatggcgacgcccgggagcactgga 4500
Db 4441 CGGCGGAGCTTCCACCCGTGCACCCGTGATGTGATGCCGACGCGCGGACCTGGA 4500
Qy 4501 acgtgctgtacagagagagcctgaagaccgcccagggacccagcagctcgggggg 4560
Db 4501 ACGTGCCTGTACGAGGAGAGCTTGAAGAGCGCGGCCAGGGACCCAGCTCGGGGGG 4560

Db 6721 TGCTGACCCGCTGCTGTACCGGGATGGGAACAGACAGCGGCCAGCAGATGGGCTCAA 6780
Qy 6781 gtctcaggcaacaccagccagccagctcttctcagcaagctgacgagagcaatc 6840
Db 6781 GTCTCCAGGCAACACAGCCAGCCGCCAGCCCTTCTTACAGAGCTGTACCGAGCAATC 6840
Qy 6841 cgcctagtgtcaagtcacgaagaagagatcaacaagaagctgaacaccccaacacccgaa 6900
Db 6841 CGCCATGTCAAGTCCAAAGACAGAGATCAACAAGAAGCTGAACACCCACACCGGAA 6900
Qy 6901 ttagcctaataataatacagcaagccttggaagagagatcttcaataatgcccgcaatc 6960
Db 6901 TGAGCCTGAATACAAATATACAGCAGCCTGGGACGAGATCTTCAATATGCCCGCATCAC 6960
Qy 6961 cgaacagccttatgacctatagaagcagcagcagcagcagcagcagcagcagcagcagc 7020
Db 6961 CGGAACAGCCTTATGACCTATAGAAGCAGCGGTGAGGAACATGCGACACCAACAT 7020
Qy 7021 ggggctggagggcctaattagaaggcactcatgggttaataatgaccagtggaagagtc 7080
Db 7021 GGGGCTGGAGGCCATAATAGAAAGCCTCATGGGTAATATGACAGTGGGAAGATC 7080
Qy 7081 cccgcctcagcgccttaatttaacccctctgtaacccctgtaacccctgtaacccctgta 7140
Db 7081 CCCGCGCTCAGCGCAATGTCTTTAACTCTGTAATGCCAGTGCAGCCTGCCCGCTGC 7140
Qy 7141 tatgccataacgcctgctgacgacgagtgaccacacactcaactcgcagctgagcag 7200
Db 7141 TATGCCATATACCGTGTGACGACGAGTGACACACACTCACCTGCCAGGTGGCGG 7200
Qy 7201 cgggaagcccaaggtctctgacagaccagcagcagcagcagcagcagcagcagcagc 7260
Db 7201 CGGAAGCCCAAGGTCTCTGGCAGACCCAGCAGCGCGAAGCCAAAGTCCCGCGCCG 7260
Qy 7261 cctggcatctgggacccgacacccctctgtctcctcagtgacactgagcagcagcagc 7320
Db 7261 CTTGGCATCTGGGACCGCGCCACCTCTGTCTCTCTCAGTCTGAGGAGGAGATGCA 7320
Qy 7321 ccgcgagcgcctccacaaacgcgtgaggagacagccctcgtcgcagcagcttcac 7380
Db 7321 CGCGGACGCGCTCACCAACCGGTGTGGAGGACAGGCCCTGCTCCGACAGGTTCAC 7380
Qy 7381 gccattccctacaccccttgatcagcagcagcagcagcagcagcagcagcagcagc 7440
Db 7381 GCCATTCCCTTACAAACCCCTGATCATCGGTGTCAGCGCGGTGTATGGCTTCCACC 7440
Qy 7441 cccacggcctcccgcgagcagcagcagcagcagcagcagcagcagcagcagcagc 7500
Db 7441 CCCACGGGCTCCCGCGGCGAGCGGCGCTCTGCTGGCCCGCCACACGCTGGGACGA 7500
Qy 7501 ggagcccaagcactgctctgctcagtagacagacactctccagacagcagtgactcag 7560
Db 7501 GGAGCCCAAGCCTGCTCTGCTCGCAGTAGAGACACTCTCCGACAGCGAGTACTCAG 7560
Qy 7561 aacagggcgggggggggcgggcggtgagcagcagcagcagcagcagcagcagcagc 7620
Db 7561 AACAGGGCGGGGGGGGGGGCGG-----TGTCAGGTCCCGAGCGAGCCACAGGAA 7610
Qy 7621 cggcctcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7680
Db 7611 CGGCCCTCAGAGGGGGGGGGGGCTGCGACATCCGCCAACCAAGGAAGAGGCCCTGAGTC 7670
Qy 7681 cgcctggcctccatccatctgctcgtccagagcagcagcagcagcagcagcagcagc 7740
Db 7671 CGCCTGCGCCTCCATCATCTGCTCGTCCAGAGCGGATCTTGGCTGTCTAAGCCTT 7730
Qy 7741 aactaagactcccgcccggtggtcctgtgacagcagcagcagcagcagcagcagcagc 7800
Db 7731 AACTAAGACTCCCGCGCGGCTGGCCCTGTGCGACACTTACTCAGGGGATGTTTACCTG 7790
Qy 7801 gtgctcggagagagagagagcagcagcagcagcagcagcagcagcagcagcagcagc 7860
Db 7791 GTGCTCGGGAAGAGGGGAAGGGGGCGGGGAGGGGCGACGGCGGTGTGGCAGCCAC 7850

Qy 7861 acacggcgccagggcgagggagcccaagcagcagcagcagcagcagcagcagcagcagc 7920
Db 7861 ACACAGGGCGCAGGGCGCCAGGGACCCAAAGCAGGATGACACGACGACCTCCAGCCAC 7910
Qy 7921 tgcctccccgaatgcatcttggaaacaaagtctaaactgagctcgcagcagcagcagcagc 7980
Db 7911 TGCTCCCGCGAATGCAATTTGGAACCAAGTCTAACTGAGCTCGCAGCCCCCGGCCCT 7970
Qy 7981 cctcgcctcccatcccgcttagcgtctggaagatggaagatggaagatggaagatggaag 8040
Db 7971 CCTCCGCTCCCATCCGCTTAGCGCTTAGCGCTGACAGATGAGCAGCGCCCTGTCCAGCC 8030
Qy 8041 ccagtgcctgctccggtcccaacagactcccaacagactcccaacagactcccaacagact 8100
Db 8031 CCAGTGCCTCTGTTCCGCTCCCGACAGACTGCCCGACGCAACAGAGATGCTGGGAACCA 8090
Qy 8101 gtcagggcagctggcgacaaaaggcgctgagcagcagcagcagcagcagcagcagcag 8160
Db 8091 CTCAGGCAGCTGGCGGACAAAAGGGCCAGTGGCGCTGGGGGACGATGCTCCGA 8150
Qy 8161 ggaactgactgttttttccacacatcgttgcgcagcagcagcagcagcagcagcagcag 8220
Db 8151 GGACTGGACTGTTTTTTTTCACACATCGTTGCCGAGCGGTGGGAAGGAAGCAGATGA 8210
Qy 8221 aatgatgtgtgtttacagaggtatattttgataccttcaatgaatgaatgaatgaat 8280
Db 8211 AATGATGTGTGTGTACAGGGTATATTTTGTATACCTTCAATGAATGAATGAATGA 8270
Qy 8281 ttacgcaagggaagacttaccagctattactgctgctgctgctgctgctgctgctgct 8340
Db 8271 TTACGCAAGGAAGGACTTTACCCAGTATTACTGCTGCTGCTGCTGCTGCTGCTGCT 8330
Qy 8341 ttcaagagcggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 8400
Db 8331 TTCAAGAGCGGTGTGCAGGGCGCAGTCGGTACCCCATCATCTCGAGGACCAAGGGG 8390
Qy 8401 ggggactgctcgtcacgccccgctgctccctccctccctccctccctccctccctccct 8460
Db 8391 GGGGACTGCTGCTCAGCGCCGCTGTGCTCCTCCCTCCCTCCCTCCCTCCCTCCCT 8450
Qy 8461 ttcatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 8520
Db 8451 TTCTGATGCTGATTTCTGTCGCGCCATTTGCGCAGGTGCTGCTGCTGCTGCTGCTGCT 8510
Qy 8521 cgtgcttctaattaaaagcgaattatactccaaaaaatactccaaaaaatactccaaaa 8571
Db 8511 CGTCGTTCTAATTAAGAGCAATTATATCTCCAAAAAATACTCCAAAAAATACTCCAAAA 8561

RESULT 2

AF125672 LOCUS AF125672 8686 bp mRNA PRI 04-APR-1999
DEFINITION Homo sapiens silencing mediator of retinoic acid and thyroid hormone receptor extended isoform (SMRTE) mRNA, complete cds.
ACCESSION AF125672
VERSION AF125672.1 GI:4559297
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 8686)
AUTHORS Park, E.J., Schroen, D.J., Yang, M., Li, H., Li, L., and Chen, J.D.
TITLE SMRTE, a silencing mediator for retinoid and thyroid hormone receptors-extended isoform that is more related to the nuclear receptor corepressor
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (7), 3519-3524 (1999)
MEDLINE 9919215
AUTHORS 2 (bases 1 to 8686)
TITLE Chen, J.D.
JOURNAL Direct Submission
TITLE Submitted (03-FEB-1999); Pharmacology and Molecular Toxicology,

QY 3337 caagcaccacagcgtctcctcgagagcgaataatagtgccatctcccaagaatgtcgttcca 3396
D 3462 CAAGCACCACCGAGTCTCGAGAGGCAAAATAGTGCCATCTCCCAAGGAATGTCCGGTCCA 3521
QY 3397 gctccacgtcccgtaactcaagagcatgccaagagccccgggtggccctgtcaccatggggct 3456
D 3522 GCTCCACGTCCTCGTACAGAGCATGCCAAGGCCCGGTGGGGCTGTACCATGGGGCT 3581
QY 3457 gccctgcccattgaccccaaaaagctggcacccttcagcggagtgaagcaggagcagct 3516
D 3582 GCCCTGCCCATGACCCCAAAAGCTGGCACCCCTTCAGCGGAGTGAAGACGAGGACGCT 3641
QY 3517 gtccccacggggccaggctggccacccggagagcgtgggggtgcccacagccccaggagcc 3576
D 3642 GTCCCCACGGGGCCAGCTGGGCCACCGAGAGGCTGGGGGTGCCACAGCCCAGGAGGC 3701
QY 3577 gtcctgctgagagggacagctctggtctgaagttccggggggaagcatcaccaaagcat 3636
D 3702 GTCCGTGCTGAGAGGGACAGCTCTGGGCTCAGTTCCGGGGCGGAAGCATCACAAAGGCAT 3761
QY 3637 tccagcacacgggtgcctcgagcagcgccatcacatcacgcgctccatcacccacgg 3696
D 3762 TCCAGGACACGGGTGCCCTCGGACAGCGCCATCACATACCGCGGCTCCCATACCCACGG 3821
QY 3697 cagccagctgacgtcctgtacaagggcaccatcacaggatcatcgcgaggagcagccc 3756
D 3822 CACGCCAGCTGACCTCTGTACAAAGGCAACCATCACAGGATCATCGCGAGGACAGCCC 3881
QY 3757 gagtctgtgaccgcggccggaggagacgctgcccgaaggccacgtctatcacaagg 3816
D 3882 GAGTCGCTTGGACCGCGCGGGAGGACAGCTGCCCCAAGGGCCACGTCATCTACGAAGG 3941
QY 3817 caaagagggcacgtctgtctatgaggtggcatgtctgtgaccagtgctccaagga 3876
D 3942 CAAGAAGGGCCACGCTGTCTGTATAGGGTGGCATGTCTGTGACCCAGTCTCCAAGGA 4001
QY 3877 ggaaggcagaagcagctcaggacccccccatgagacggcgcccccgaagcgacactatga 3936
D 4002 GGACGGCAGAAAGCAGCTCAGGAACCCCCCATGAGACGGCGCCCCCAAGCGCACCTATGA 4061
QY 3937 catgatggaggggccgctgagagagcattctctcagcagcatcgaaggctcatggg 3996
D 4062 CATGATGGAGGGCCGCTGGGCAAGAGCATCTCTCAGCCAGCATCGAAGGCTCATGGG 4121
QY 3997 ccgtgcatcccgccgagcgacacagccccccacactcaaaagcagcacacacatccg 4056
D 4122 CCGTGCCATCCCGCGGAGGACACAGCCCCCACCCTCAAGAGCAGCACCATCCG 4181
QY 4057 cgggtccatcacacaaggatccctcggtctctacgtgagggcagagagactacctcg 4116
D 4182 CGGTCTCATCACAAAGGATCCCTCGTCTACGTGAGGSCACAGAGGACTACCTGG 4241
QY 4117 tcggggagcgaagctctaaagggggggagcagcctccgccccccacgcgcctcagggga 4176
D 4242 TCGGGAGGCCAAGCTCTAAAGCGGGAGGACGCGCTCCGCCCCCAGCGCCCTCACGGGA 4301
QY 4177 cctgacagagcctacaagcagcagccctggccccctaaagctgaagcggcccatga 4236
D 4302 CCTGACCGAGGCTTACAAGACGACAGGCCCTTGGGCCCTGAAGCTGAAGCGCGGCCATGA 4361
QY 4237 ggggctggtgcccagctgaaggaggcggggcgctctccatccatgagatcccgcgcgagga 4296
D 4362 GGGCTGTTGGCCACAGTGAAGGAGGCGGGCGCTCCATCCATGATCCGCGCGGAGGA 4421
QY 4297 gctcgccacacgcccagctgccccctggccccggccgctcgaaggagggctccatcac 4356
D 4422 GCTCGGCACACGCGCGAGCTGCCCTTGGCCCGCGCGGCTCAAGGAGGGCTCCATCAC 4481
QY 4357 qcaggcaccctcgaatcacacacgcgcgctccaccactgctccaaaagacga 4416
D 4482 GCAGGACCCCGCTCAAGTACGACACCGCGCGCTCCACCCTGGCTCCCAAAAAGCACGA 4541

QY 4417 cgtacgtccctctatcgcgagcagccccgcggagcttccccaccgtgacccccgtggatgt 4476
D 4542 CGTACGCTCCCTCATCGCAGCCCGCGGCGAGCTTCCACCCTGTCACCCGCTGGATGT 4601
QY 4477 gatggcgaagccccgggcaactggaaactgctgtacagagagagcctgaagagccggcc 4536
D 4602 GATGGCGAGACCCCGGGCAGCTGGAAAGCTGCTGTACGAGGAGAGCTGAAGAGCCGCC 4661
QY 4537 agggacccgcagcagctcggggggtccattgcgcgcgcccggtcatgtgtcctga 4596
D 4662 AGGACCGCCAGAGCTCGGGGGCTCCATTGCGCGCGCGCCCGGTCATTGTGCTCTGA 4721
QY 4597 gctgggtaagccgcggcagagccccctgacctatgaggaccacggggcacccctttgcgg 4656
D 4722 GCTGGTAAAGCCGCGGAGAGCCCTGACCTATGAGGACACAGGGGCACTCTGCGCGG 4781
QY 4657 ccactccccagaggttcgcccgtgacatgctggggagccccgcgcgctgcaggaagg 4716
D 4782 CCACCTCCACAGAGTTTCGCCCTGACCATCGGGAGCCCCACGCGCCTGCAGGAGGG 4841
QY 4717 cagcttctgtccagcaagcgtcccagagcagaaagctgagtcgacgctcgtgagat 4776
D 4842 CAGCTTTCGTCACGAAGGATCCAGGACGAAAGCTGAGTCGACGCGCTCGTGAGAT 4901
QY 4777 cgcaagtcccccgcacagcaccgtgcccgcgagcacccacacaccccatctcgccctatga 4836
D 4902 CGCAAGTCCCGCACAGCAGCCGTGCCGAGCACCACACACCCCATCTCGCCCTATGA 4961
QY 4837 gcaactgcttggggctgagtgctggctgagactgtatcgagccacatcccccttggcctt 4896
D 4962 GCACCTGCTTGGGGCGGTGAGTGGCTGGACCTGTATCGACGCCACATCCGCCCTGGCCT 5021
QY 4897 cgacccacactccatccccgcgcgcatccctctgagcagcagccctgctactactacc 4956
D 5022 CGACCCACCTTCCATACCCCCCGGCATCCCTCTGACGACGCGCTGCTACTCTGCC 5081
QY 4957 ccgacactggcccccaacccactaccgcgacactgtaccacacccactcactccgcgg 5016
D 5082 CCGACACCTGGCCCCCAACCCACCTACCCGACCTGTGTACCCACCTACCTCATCCGCGG 5141
QY 5017 ctaccccgacagcgcgctggagaaacccgcgagacacacacacacacacacacacac 5076
D 5142 CTACCCCGACAGCGCGCGCTGGAGAACCCGCGAGACCATCATCAATGACTACATCACCTC 5201
QY 5077 qcagcagatgcaccacacacgcgcaccccatgcccagcgagctgatatgctgagggg 5136
D 5202 GCACGATGACACACACACGCGCCACCGCCATGGCCCGCAGCGAGCTGATATGCTGAGGG 5261
QY 5137 cctctcgcccgcgagtcctcgctggcactcaactacgctcggtgcccccgaggcatcat 5196
D 5262 CCTCTGCGCGCGAGTCTCGTGGCACTCAACTAGCTGCGGTCCCGGAGGCTCAT 5321
QY 5197 cgactgtcccaagtgcacacactgctgtgctgctgctgctgctgctgctgctgctgctg 5256
D 5322 CGACTGTCTCCCAAGTGGCCACACTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 5381
QY 5257 caccgcatgacgcgcttgcctacctccccacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5316
D 5382 CACGCCATGACCGCGCTTGGCTTACTTCCACCGCGCCCGCAGCCCTTCACGAGCGCGCA 5441
QY 5317 cagcagctccccactctccccagaggtcccaacacttgacaaaacacacacacacacac 5376
D 5442 CAGCAGTCCCTCTCTCCCGAGGAGTCCAACACACTTGACAAAAACACACACACCTC 5501
QY 5377 ctgctcgagcgagcgagacccggggtcgagagcggggacccggggtcgggagcgggaaaa 5436
D 5502 CTCCTCGAGCGGGAGGAGACCGGGATCGAGAGCGGAGCGGGATCGGGAGCGGGGAAA 5561
QY 5437 gtcctactcactgctccaccacgcgctggagcagcaccatctgagagacctgtacaga 5496
D 5562 GTCCATCTTCACTCCACCACGAGCGGTGGAGACGACGCCCTATCTGGAGACCTGGTACAGA 5621
QY 5497 gcagagcagcggcagcagcgagcggcggggtgggggggagcagcagcgccccgc 5556

|||||
Db 5622 GCAGAGACGCGCAGCAGCGGAGCGGGGGTGGGGCAGCAGCAGCGCGCCCGC 5681
QY ctccactcccatgccacacagcaactgcgccatctccctcggaccacagatgcctcca 5616
Db 5682 CTCACCTCCATGCCATGCCACACAGCAGCTCCGCCATCTCCCTCGAACCGATGCCCTCCA 5741
QY gcagagacccagtgtgtctcaacacacagggcatgaaggggtatcatcacgcgtgtggagcc 5676
Db 5742 GCAGAGACCCAGTGTCTTACACACAGGATGAGGGGTATCATCACCGTGTGGAGCC 5801
QY cagcaagccacggtctctaggtccacatccactctcaccctgctccacagctgcac 5736
Db 5802 CAGCAAGCCACCGTCTGAGGTCCACCTCCACCTCCCTCCACCGTTCGCCAGCTGCCAC 5861
QY attccacactgccacacactgccacactggcggaacctcgtatgggtctacacctacct 5796
Db 5862 ATTCCCACTGCCACCTGCCACTGGGGGACCTCCGATGGGGTCTACCTTACCT 5921
QY catggagccctcttctgtcccaagaggccccccgggtgcgcgggccagagcgccctcg 5856
Db 5922 CATGGAGCCGTCITGCTGCCAAGAGGCCGCCCGGTGCCCGCCAGAGCGGCCCGC 5981
QY agcagacaccgccaatgcttctctctcccaagccccccagccccgctccgggtggagccgc 5916
Db 5982 AGCAGACACCGGCCATGCTCTCTGCCAAGCCCGCAGCCCGCTCGGGCTGGAGCCCGC 6041
QY ctctctccccagcaaggtcggagccccccctctctctctctctctctctctctctctc 5976
Db 6042 CTCCTCCCCAGCAGAGGCTCGGAGCCCCCGCCCTAGTGTCTCTCTCTCTCTCTCTCT 6101
QY caccatcgcccgcaacccctcgaaagaaactcgcaactcaccacgcagcccgagcccgcc 6036
Db 6102 CACCATCGCCGCGACCTCTGCGAAGAACTCTGCACCTCAACCAGCCAGCCCGGCCCGC 6161
QY ggcgcaactcct 6096
Db 6162 GCGCGCACTGCTCGGCTCGGACCCCGCAGCCGCGGAAAGAACTCAAAAGTAAACCTTTTC 6221
QY catcagaaactgaactccgtctctctgtttacacggagcaactacagcccgcaag 6156
Db 6222 CATCCAGAACTGGAACTCCGTTCTCTGGGTTCACCGGCGAGCAGCTACAGCCCGGAAG 6281
QY ggtgagccctcgacccctgtgactcaccagctctgaccacagcaaggggtcccca 6216
Db 6282 GGTGAGCCCGCTCAGCCCTGTGAGCTACCCAGTCTGACCCAGCAGAGGGCTCCCCAA 6341
QY gcaactggaagactcgacaagaccactggaggggagctgaggcccaagcgccag 6276
Db 6342 GCACGTGAAGAGTTCGAACAAGACCACTGTGAGGGGAGCTGGCGGCCAAGACGCCAGG 6401
QY cccggtgaagcttgaggggagggccgcccactccacacactcgccgctgacctgag 6336
Db 6402 CCCGTGAAGCTTGGCGGGAGGGCGCCCACTCCACACCTGGCGGCTGCTCTGAGAG 6461
QY ccagccctctccagccgctgtccagacgcccaggggtcaaaagtaccagcggt 6396
Db 6462 CCAGCCCTCTCCAGCCCGCTCTCCAGACCGCCAGGGGTCAAAAGTCAACAGCGGGT 6521
QY ggtcaccctgcccagacatcagtgaggtatcaacagagactacacccggcaacccc 6456
Db 6522 GGTCAACCTTGCCCGCAGCATCAGTGAAGTTCATCACAGAGACTACACCCGGCACCAACC 6581
QY acagcagctagcacacctgcccgcctctactctctctctctctctctctctctctctct 6516
Db 6582 ACAGCAGCTCAGCGACCCCTGCGCGCCCGCCCTCTACTCTCTCTCTCTCTCTCTCTCT 6641
QY cgtctggaactccgcccacacagtgacctctactctccgcccccgagaccatgtgc 6576
Db 6642 CGTCTGGACCTCGCGCCCGCAGCAGTACCTCTACTCTCTCTCTCTCTCTCTCTCTCT 6701
QY cccggccgctgggtccccccacagcggaagggggcaaggtctccagagccaaacagac 6636
|||||

Db 6702 CCCGCGCGTGGCTCCCCCACAGCGAAGGGGGCAAGAGGTCTCCAGAGCCAAACAAGAC 6761
QY gtcggtcttggttggttggtgaagagcggtattgaacctgtgtccccacccggagggcatgac 6696
Db 6762 GTCGCTCTTGGGTGGTGGTGGAGACGGTATTGAACCTGTGTCCCCACCGGAGGCGATGAC 6821
QY gtagccagggcactccccggagtgctgttacccgctgtgtaccggggtggtgggaacagac 6756
Db 6822 GGAGCCAGGGCAGTCCCGGAGTGTGTGTACCCGCTGTGTGATCCGGGATGGGGACAGAC 6881
QY gtagccagcaggtaggtgtccaaagtctccaggcaacaccagccagccgagcctcttt 6816
Db 6882 GGAGCCAGCAGGATGGCTCCAAAGTCTCCAGGCAACACACAGCCAGCCGCGCAGCTTCTT 6941
QY cagaagctgaccagagagcaactccgccaatgtcaagtccaagaagcaagatcaacaa 6876
Db 6942 CAGCAAGCTGAGCCAGAGAACTCCGCCATGGTCAAGTCCAAAGAAAGCAAGAGATCAACAA 7001
QY gaagctgaacacccac 6936
Db 7002 GAAGCTGAACACCCAC 7061
QY gatctcaatatgcccgcatacccggaacaggccttatgacctatagaagccagggcgt 6996
Db 7062 GATCTTCAATATGCCCGCCATCACCGGAACAGGCTTATGACCTATAGAACCGAGCGGT 7121
QY gcagaaacatgccagcaccac 7056
Db 7122 GCAGGACATGCCAGACCAACATGGGGCTGGAGGCCATTAATTTAGAAAGGACACTCATGGG 7181
QY taaatatgaccagtggaagagtgcccgcgctcagcgcccaatgcttttaacccctctgaa 7116
Db 7182 TAATATGACCAAGTGGGAAGAGTCCCGCCGCTCAGCGCCAAATGCTTTTAAACCTCTGAA 7241
QY tgccagtgccagcctgcccgtgtatgcccataacccgctgtgacgagcagagtgacca 7176
Db 7242 TGCCAGTGCACAGCTGCCCGCTGTGTATGCCCATAAACCGCTGTGACGACGAGGTGACCA 7301
QY cactcactcgcagagtgccgggaggaagcacaaggtctctgacagaccagcagcgg 7236
Db 7302 CACACTCACTTCGCCAGGTGGCGGGAGGCCAAGGTCTCTGACAGACCCAGCAGCGG 7361
QY aaaaagcaagtgcccgcgcccggctgggaatctgggagccggccacccctctctctc 7296
Db 7362 AAAAGCCAAAGTCCCGCGCCCGGCGCTGGCATCTGGGAGCCAGCCACCTCTGTCTCTCTC 7421
QY agtcaactcgagggagagactgcaaccccgccgagcgcctcaccacacccgctgtggagga 7356
Db 7422 AGTGCACTCGGAGGGAGACTGCAACCCCGGAGCGCGCTCAACCAACCGCTGTGGGAGGA 7481
QY cagccctctcccgaggttccacgccaattccctcaacccctgacatgagctgca 7416
Db 7482 CAGCCCTCTGTCGCGAGGTTCCAGCCATTCCTCTACACCCCTGATCATGCGGCTGCA 7541
QY ggcgggtgtcatggtctccccacccacccggcctcccgccgggagcgggcccctcgc 7476
Db 7542 GCGGGGTGTCATGCTTCCCCACCCACCGGCTCCCGCGGCGAGCGGCGCCCTCGC 7601
QY tggccccacacacgcttgagcagagggcccaagcactgctctgtcgtcagatcagagac 7536
Db 7602 TGGCCCCCACACGCCCTGGGACGAGAGGCCCAAGCCACTGTCTGTCTGCGCAGTACGAGAC 7661
QY actctccagcagcagtgactcaaaacagggcggggggggggggggggggggggggggggggg 7596
Db 7662 ACTCTCGACAGCGAGTGTCTCAAGACAGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 7711
QY tcaggtccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7656
Db 7712 TCAGGTCCCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 7771
QY aaccaaggaaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7716
Db 7772 AACCAAGGAAGAGGAGCCCTGAGTCCGCTGCGCTTCCATCCATCTGCTCCGTCAGAGCGCG 7831

Qy	7717	gcattctgtctgtctctaaagcccttaactaagaactcccccccccggcttggccctgtgcaga	7776
Db	7832	GCATCTTGGCTGTCTTAAAGCCTTAACTAAGACTCCCGCCCGGCTGGCCCTGTGCAGA	7891
Qy	7777	ccctactcagggaatgtttacctgtgtctcgggaaggaggaaggccgggagggg	7836
Db	7892	CCTTACTCAGGGGATGTTTACCTGTGTCTCGGGAAGGAGGGGAAGGGCCGGGAGGGG	7951
Qy	7837	gcacgcgagcgctgtgcagccacacacagcgcgccagggcgcgccaggacccaaacag	7896
Db	7952	GCACGCGAGCGTGTGGCAGCCACACACAGGGCGGCAGGGCGGCCAGGACCAAAACGAG	8011
Qy	7897	gatgaccagcaactccacgcccactgcctccccggatgtcatttggaaaccaaagtctaaa	7956
Db	8012	GATGACACGCACCTCCAGCCACTGCCTCCCGCCGAATGCTTTGGAACCAAGTCTAAA	8071
Qy	7957	ctgagctgcagccccccgcgcctccctccgcctcccatacccgctttagcgtcttgacaag	8016
Db	8072	CTGAGCTCGAGCCCCCGCGCCTCCCTCCGCGCTCCCAATCCGCTTAGCGCTCTGGACAG	8131
Qy	8017	atggacgcagccctgtccagccccagctgcgctcggttcccggtccccacagactgcacca	8076
Db	8132	ATGGACGCAGGCCCTGTCCAGCCCCCAGTGCCTCGTTCGGTCCGCCACAGACTGCCCCA	8191
Qy	8077	gccaaacgagattgtgtgaaaccaaactcagggccaggtggcgcgacaaaggccaggtgcg	8136
Db	8192	GCCAAACGAGATTGCTGAAACCAAGTCAGGCCAGGTGGCGGACAAAGGCCAGGTGCG	8251
Qy	8137	gcctggggggaacgagatgtcgcgagactggaactgttttttttccacacatogttgcgcag	8196
Db	8252	GCCTGGGGGGAACGGATGCTCCGAGGACTGGGACTGTTTTTTTTCACATCGTTGCCGAG	8311
Qy	8197	cggtgggaaggaaagcagatgaaatgatgtgttggtttcacaggggtatatctttgatac	8256
Db	8312	CGGTGGGAAGAAAGCCAGATGTAATGATGTGTGTGTTTACAGGGGTATATTTTGTATAC	8371
Qy	8257	cttcaatgaattaatcagatgttttacgaaaggagacttaccagttactactgtctgc	8316
Db	8372	CTTCAATGAATTAATTCAGATGTTTACGCAAGGAAGACTTACCAGTATTACTGCTGC	8431
Qy	8317	tgtgcttttgatctctgtctacacgttcaagaggcgctgcagggccgacagtcggtgacc	8376
Db	8432	TGTGCTTTTGATCTCTGCTTACCGTTCAAGAGCGCTGTGCAGGCCGACAGTCTGGTGACC	8491
Qy	8377	catcactccagagaccaaaggcgaggagactgctcgtcaccccgctgtgtctccctc	8436
Db	8492	CATCCTCGGAGACCAAGGGCGGGGAGCTCTCGTCACGCCCGCTGTCTCTCCCTC	8551
Qy	8437	ctccctctcttgggcagaatcgaatcgaatcgtattctgtggcgccatttgcgcaggg	8496
Db	8552	CTCCCTTCTTGGGCAGAAATGAATTCGATCGTATTCTGTGGCGGCCATTTGGCGAGG	8611
Qy	8497	tgtgttattctgtcatattcacacgctgttctataataaaagcgaattatactccaaaa	8556
Db	8612	TGTTGTATTCTGTCTATTTACACAGTCGTTCTAATTAATAAGGGAATATATCTCCAAA	8671
Qy	8557	aaaaaaaaaaaaaa 8571	
Db	8672	AAAAAAAAAAAAAAAA 8686	
RESULT 3			
LOCUS	HSU37146		
DEFINITION	HSU37146 5989 bp mRNA	PRI	31-OCT-1995
ACCESSION	Human silencing mediator of retinoid and thyroid hormone action (SMRT) mRNA, complete cds.		
VERSION	U37146		
KEYWORDS	U37146.1 GI:1045654		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom		

Db 2787 AAAACAAACCACCGTCTCTCGAGCGGAGCGAGACCGGGATCGAGAGCGGGACCG 2846
Qy 5419 ggtcggagcgggaaagtccatctcactcacgtccaccacgagcgtgagcagcaccccat 5478
Db 2847 GGATCGGAGCGGGAAAGTCCATCTCTCAGTCCACACGAGCGGTGGAGCAGCGACCCAT 2906
Qy 5479 ctggagacctggtacagagcagcagcggcagcagcggcagcagcggcggggtggggg 5538
Db 2907 CTGGAGACCTGGTACAGAGCAGAGCAGCGGACGAGCGGACGAGCGCGGGGTGGGG 2966
Qy 5539 cagcagcagcggccccctcccactccatgccaccacagcactgcacatccctcccg 5598
Db 2967 CAGCAGAGCGCGCGCGCTCCCACTCCCACTGCCCACTGCCCACTGCCCACTGCCCTCG 3026
Qy 5599 gacccagatgacctccagcagagacccagtgcttcacacacagcagcatgaagggtat 5658
Db 3027 GACCCAGATGCCCTCCAGCAGAGACCCAGTGTCTTCAACACAGCAGCATGAAGGTAT 3086
Qy 5659 cateacgctgtgagccagcaagccacggtcctgaggtccacctccacctccacc 5718
Db 3087 CATCACCGCTGGAGCGCAGCAAGCCACCGTCTGAGGTCCACCTCCACTCCCTCACC 3146
Qy 5719 cgttcgcccagctccacattccacatgccaccatgccaccatgccaccatggcggcaccctcga 5778
Db 3147 GCTTCGCCACGTGCCCATTTCCCACTGCCACCCACTGCCCACTGGCGGACCCCTCGA 3206
Qy 5779 tgggtctaccctaccctcatggagccgctctgtgtcccaagagagcccccgggtgcg 5838
Db 3207 TGGGTCTACCTCATCTGAGCGCGTCTTGTGCGCCAGGAGGCCCCCGGGTGC 3266
Qy 5839 ccggccagagcggccccgagcagacaccggccatgctctctcgcgaagcccccagcccg 5898
Db 3267 CCGGCAGAGCGGCGCGGAGCAGACACCGGCGCATGCCCTCTCGCAAGCGCCCGCAGCCG 3326
Qy 5899 ctccgggtcagcccgctctcccccagcaaggctcagagcccgcccgccctagtacc 5958
Db 3327 CTCGGGCTGAGCGCGGCTCTCTCCCCAGCAAGGGCTCGGAGCCCCGGCCCTAGTGCC 3386
Qy 5959 tctgtctctgcccagccacatagcccgaccctcgcgaagaacctcgacacctcaaca 6018
Db 3387 TCCCTGCTCTGGCCAGCCACCATCGCCCGCACCCCTCGGAAGAACCTTCGACCTCACCA 3446
Qy 6019 cgcagcccgagcccgccggcgcacactgcctcggcctcgcagcccgccagccgggaaagac 6078
Db 3447 CGCCAGCGGACCGCGCGGCGCACCTGCGCTCGGCGCTCGGACCGCCGACCGGGAAAGAC 3506
Qy 6079 tcaagtataacccctttccatccaggaactgaaactccgtctctggtttaccacggcag 6138
Db 3507 TCAAGTAAACCTTTTCCATCCAGGAACCTGGAATCCCGTTCTTGGGTATCCACGGCAG 3566
Qy 6139 cagctacagcccggaagggtggagccgctcagccctgtgagctacccagctctgaacca 6198
Db 3567 CAGCTACAGCCCGGAAGGGGTGAGCCCGTCAGCCCTGTGAGCTACCCAGTCTGACCCA 3626
Qy 6199 cgacaagggttccccagcaccttgaagagctcgacagagccacctggaggggagct 6258
Db 3627 CGACAAGGGGTTCCTCAAGCACCTTGAAGAGCTCGACAAGAGCCACCTTGGAGGGGAGCT 3686
Qy 6259 gcggcccaagcagcagcccggtgaagcttggcgggagagcccccacctcccaacct 6318
Db 3687 GCGGCCCAAGCAGCGAGCCCGGTGAAGCTTGGGGGAGGCGCCCGCCACCTTCCACACT 3746
Qy 6319 gcggccgctgctgagagccagcctcgtccagccgctgctccagagccgcccaggggt 6378
Db 3747 GCGGCCGTGCTGAGAGCCAGCCCTCTCTCAGCCCGCTGTCTCAGACCGCCCGCCAGGGGT 3806
Qy 6379 caaagggtaccaggggtgttcaacctggcccagacacatcagtgaggtcatcacagga 6438
Db 3807 CAAGGTCAACAGCGGGTGTCACTTGGCCCGACACATCATCAGTGAGGTCTACACAGGA 3866
Qy 6439 ctacacccggcaccacccacagcagctcagcagccctgcgcccctctctactcctt 6498
Db 3867 CTACACCGGACCAACCCAGCAGCTCAGCGCACCCCTGCGCCGCCCGCCCTCTACTCTT 3926

Qy 6499 ccctggggccagctcccgctcctggaactccgcccgcgcccaccagtgacctctacctcc 6558
Db 3927 CCTTGGGGCCAGTCCCGCTCTGGACCTCGCGCGCCACCCAGCTGACCTTACTCTCC 3986
Qy 6559 gcccccggaccatggtgcccgccgctggcttccccccacagcgaagggggcaaggtc 6618
Db 3987 GCGCCCGGACCATGTGTGCCCGCGCTGGCTCCCGCCACAGCAAGGGGGCAAGAGTTC 4046
Qy 6619 tccagagcacaagaacgctcggtcttgggtggtggtgagcaggtattgaacctgttc 6678
Db 4047 TCCAGAGCCAAACAAGACGTGGTCTTGGTGGTGGTAGGACGGTATTGAACCTGTGTTC 4106
Qy 6679 ccacccggagggcatgacggagccaggtccctcccgagtgctgtgtacccgctcgtca 6738
Db 4107 CCCACCGGAGGCGATGACGGAGCCAGGCGACTCCCGGAGTGTGTGTACCCGCTGCTGTA 4166
Qy 6739 ccggatgggaacagacggagcccgagcaggtgggttccaaagtctccaggcaaacag 6798
Db 4167 CCGGATGGGAACAGAGCGGAGCCAGGAGTGGCTCCAAAGTCTCCAGGCAACACCAG 4226
Qy 6799 ccagcccgaccccttctcagcaagctaccagagacactccgcatggtcaagtccaa 6858
Db 4227 CCAGCCCGCAGCTTCTTACGAAGCTACCGAGAGCAACTCCGCCATGGTCAAGTCCAA 4286
Qy 6859 gaagcaagagatcaacaagaagtgaacacccacacccggaatgagcctgaatacaatat 6918
Db 4287 GAAGCAAGAGATCAACAAGAAGCTGAACACCCACCAACCGAATGAGCCTGAATAAATAT 4346
Qy 6919 cagccagcctgggacggagatttcaatatgcccgcctaccacggaaagccttatgac 6978
Db 4347 CAGCAGCGCTGGAGCGGAGATCTTCAATATGCCCGCCCATCACCGAAGAGGCTTATGAC 4406
Qy 6979 ctatagaagcagcggtgcaggaacatgccagcaccacatgggctggagggcataat 7038
Db 4407 CTATAGAAGCCAGGCGGTGCAGAAACATGCCAGCACCAACATATGGGCTGGAGGCCATAAT 4466
Qy 7039 tagaagcactcatgggttaatatgacagtggaagagtcgcccgccgctcagcgcaaa 7098
Db 4467 TAGAAAGCACTCATGGGTAAATATGACAGTGGGAAGAGTCCCCGCCCTCAGCGCCAA 4526
Qy 7099 tgcctttaacctctgaatgccagtcgagcctgcctgctatgccaataaacgctgc 7158
Db 4527 TGTCTTTAACCTCTGAATGCCAGTCCAGCTGCCAGCTGCCGCTGTATGCCATTAACGCTGC 4586
Qy 7159 tgacggagcagtgaccacacactcactcgcagtgagcggcgggaaagcgaagcttc 7218
Db 4587 TGACGACGGAGTGACACACACTCACTTCGCGAGGTGGCGGGGAAAGGCAAGTCTC 4646
Qy 7219 tggcagacccagcagcgaagcaagtcgcccgcccgccgctggcaatctgggagccg 7278
Db 4647 TGGCAGACCCAGCAGCCGAAAGCAAGTCCCGCGCCCGCGGCTGGCATCTGGGAGCG 4706
Qy 7279 gccacctctgtcctcactcagtcgactcggagggagactgcaaccccgccagcgcctcac 7338
Db 4707 GCCACCTCTGTCTCTCAGTGCACCTCGAGGAGAGCTGCAACCCCGCGAGCCTCGCTAC 4766
Qy 7339 caacgcgtgtggagagacagccctcgtccgaggttccagccattccctacaacc 7398
Db 4767 CAACCGCTGTGGGAGGACAGGCCCTCTGTCCGACAGGTTCACGCCATTCCCCCTACAACC 4826
Qy 7399 cctgatcatgvgctgcaggcggtgtctatggttccccacccccccaggggctccccgc 7458
Db 4827 CCTGATCATGGGCTGACAGCGGGTGTCTATGGCTTCCCCACCCCGCGGCTCCCGC 4886
Qy 7459 gggcagcgggccccctcgtgcccccccaacacgctcgggagcagggcccaagcactgct 7518
Db 4887 GGGCAGCGGGCCCCCTCGCTGGCCCCCACACGCTGGGAGGAGGAGCCCAAGCCACTGCT 4946
Qy 7519 ctgctcgcagtagcagacactctccgacagcaggtgactcagaacagggcgggggggc 7578
Db 4947 CTGCTCGCAGTACGAGACACTCTCCGACAGGAGTGTACTAGAACAGGCGGGGGGGC 5006

PDIGTTPARSPHSEGGKSPBSKTSVLGSEDAIEPVSPPEGMTGPHARSAYVPLL
YRDEGEPFMGSKPGNTSPQAPFKLTENSAWVKKEILNKLTNHNDEEY
NIOGOTFENIPATIGACLTGTSQAOVHAESTNMGLEITRKALMGKYDNEEPP
GLASANTPLNASLPAAMPTTADGRSDHIALTS PGCGGKAKVSGRPSRKKASAP
SPPPGLAAGSPGLAGPHHAEDEEPPKLCQSYETLSDSE*
SPPPGLAAGSPGLAGPHHAEDEEPPKLCQSYETLSDSE*

	BASE COUNT	ORIGIN
1	0.000	0.000
2	0.000	0.000
3	0.000	0.000
4	0.000	0.000
5	0.000	0.000
6	0.000	0.000
7	0.000	0.000
8	0.000	0.000
9	0.000	0.000
10	0.000	0.000
11	0.000	0.000
12	0.000	0.000
13	0.000	0.000
14	0.000	0.000
15	0.000	0.000
16	0.000	0.000
17	0.000	0.000
18	0.000	0.000
19	0.000	0.000
20	0.000	0.000
21	0.000	0.000
22	0.000	0.000
23	0.000	0.000
24	0.000	0.000
25	0.000	0.000
26	0.000	0.000
27	0.000	0.000
28	0.000	0.000
29	0.000	0.000
30	0.000	0.000
31	0.000	0.000
32	0.000	0.000
33	0.000	0.000
34	0.000	0.000
35	0.000	0.000
36	0.000	0.000
37	0.000	0.000
38	0.000	0.000
39	0.000	0.000
40	0.000	0.000
41	0.000	0.000
42	0.000	0.000
43	0.000	0.000
44	0.000	0.000
45	0.000	0.000
46	0.000	0.000
47	0.000	0.000
48	0.000	0.000
49	0.000	0.000
50	0.000	0.000
51	0.000	0.000
52	0.000	0.000
53	0.000	0.000
54	0.000	0.000
55	0.000	0.000
56	0.000	0.000
57	0.000	0.000
58	0.000	0.000
59	0.000	0.000
60	0.000	0.000
61	0.000	0.000
62	0.000	0.000
63	0.000	0.000
64	0.000	0.000
65	0.000	0.000
66	0.000	0.000
67	0.000	0.000
68	0.000	0.000
69	0.000	0.000
70	0.000	0.000
71	0.000	0.000
72	0.000	0.000
73	0.000	0.000
74	0.000	0.000
75	0.000	0.000
76	0.000	0.000
77	0.000	0.000
78	0.000	0.000
79	0.000	0.000
80	0.000	0.000
81	0.000	0.000
82	0.000	0.000
83	0.000	0.000
84	0.000	0.000
85	0.000	0.000
86	0.000	0.000
87	0.000	0.000
88	0.000	0.000
89	0.000	0.000
90	0.000	0.000
91	0.000	0.000
92	0.000	0.000
93	0.000	0.000
94	0.000	0.000
95	0.000	0.000
96	0.000	0.000
97	0.000	0.000
98	0.000	0.000
99	0.000	0.000
100	0.000	0.000

Query Match 58.8%; Score 5037.2; DB 94; Length 8544;
Best Local Similarity 78.3%; Pred. No. 0;
Matches 6770; Conservative 0; Mismatches 1493; Indels 387;

Qy	1	catgtcgggtccacacagctgtgacacagacgtgagggccactgagccccgtacc	60
Db	159	CATGTAGGATCCACACAGCCTGTGGCACAGACATGCGGGCTGCTGAGCCCGCTACCC	218
Qy	61	gccccacagcctttaccacagtgagatcgccgagacgacacagagctgggctct	120
Db	219	ACCCATGGCATCTCTACCCGGTGAGATAGGCCGGTCCACACAGGAGTGGGGTGCT	278
Qy	121	ggagtaccagcacactcccgactatgctctcccactgtgcgggtcccatatcca	180
Db	279	TGAGTACCAACACCCCGTGACTACACCTCACACCTGTACCCGGTTCATCATCCA	338
Qy	181	gccccaggggagggcctcctgtgtgtgattccagccgggaatgaacggtccca	240
Db	339	GCCACAGAGGAGGGCGGCTCACTGTGTGAGAGTTCCAGCCTGGGAGTGAACGGTCTCA	398
Qy	241	ggagctccactgggcccagagtcacactatacctgcgcgagctggggaagtcaagat	300
Db	399	GGAGCTCCACTCGCGCCTGAGTCCCGCACGTTCTGCTGAGCTGGGAAGCCCGACAT	458
Qy	301	ggagttcattgaagcaagcgccctcgcttagagctgctctgacccctctcgacc	360
Db	459	AGAA TTCACGACGAGCAGGCGCCCGCTGGAGCTACTACCCGATACCTGTCTGGCC	518
Qy	361	gtcacccctgtggccacggccagcctgcggatctgaagacctcacaaaggaccgtag	420
Db	519	ATCACCCCTGCTGCCACTGGGACGCGAGTGGGTCTGAAGACCTTACCAAGGACCGTAG	578
Qy	421	cctgacgggcaagctggaaacgggtgtctccccccagccccccgcacactgacctgagct	480
Db	579	CCTGGCAGGCAAGCTGGAGCCTGTGTCACTCCCTCCAGTCCCGCGCAGCGCTGACCTGAGCT	638
Qy	481	ggagctgtgtccgcacggctgccaaagagagctgattccagaaca---tgagccgct	537
Db	639	AGAGCTGGCCCATCTGCACTGTCTAAGGAGGAGCTGATCCGAACAGANTGGACCGCT	698
Qy	538	ggaccgagatcacccatggtagagcagcagatctcttaagctgaagaagaagcagcaaca	597
Db	699	GGACCGTGAGATCACCATGCTAGACGACGAGATCTCCAAGCTGAAGAAGACGACGCAACA	758
Qy	598	gctggaggagggtgcgaagccgcccagcctgagagccogtgtcaccccgcccat	657
Db	759	GTTGGAGGAGGCGCCCAAGCGCGCCCAACCCGAGAGCGCTGTCTGCCACACCCCAT	818
Qy	658	cgagtgcgaagcacccgagcctggtgcagatcatctacgacgagaaacccgaaagaaggtga	717
Db	819	AGAATCAAAACCCCGAAGCCTGTGTCCAGATCATCTACGATGAGAACCGGAAGAACGCCGA	878
Qy	718	agctgcacatcggatctggaaggcctggggcccagggtggagctgccctgtacaacca	777
Db	879	AGCGGCACACGGATCCTAGAAAGCCTGGGGCCCCAGGTGGAGTGCCTTCTGTACAAACCA	938
Qy	778	gcoctcgcacaccggcgagtcatctgagaacatcaaaataaaaccaggcgcatcggaagaa	837
Db	939	GCGCTGTGACACACGCCAGTACCATTGAAAAACATCAAAATAAACCCAGCGCATGCGGAAGAA	998
Qy	838	gctaaattgtactcaagaggagggaatacacgctcggaacaataatggaagcagaagtctgt	897
Db	999	GCTGATCTTGTACTTTAAGCGGAGGAACACGCGCAAGCAGTGGGAACAGCGTCTTG	1051

Qy	898	ccagcgctatgacagctcatgagggcccttgaaaaaaggctgagcgcatcgaatacaaa	957
Db	1059	CCAGCGCTATGACCACTCATGAGCGCTGGGAGAGAGAGGTAGAGCGCATAGAGAACA	1118
Qy	958	cccgccgcgggcccaaggagacgaagtgcgcagctactacgaagaacagtttccctga	1017
Db	1119	TCCCGGAAGGAGGCCAAGGAGACCAAGGTGAGGAGTACTACGAGAAACAAGTTCCCGGA	1178
Qy	1018	gatccgcaagcagcgagctgcaggagcgcatgcagagcaggggtggggccagcgggcag	1077
Db	1179	GATCCGCAAGCAGCGGAGCTGCAGGAGCGCATGCAAGCAGAGGTGGCGCAGCGTGGCAG	1238
Qy	1078	tgggcttcctatcgccgcgcgcagcagcagcagcagcagcagcagcagcagcagcagc	1137
Db	1239	TGGGCTCTCCATGTGCGCTGCCCGCAGTGAGCATGAGGTTCTTGAGATCATTGATGGCT	1298
Qy	1138	ctcagagcaggagaacctggagaagcagatgcgcagctggcgtgatcccgcccatgct	1197
Db	1299	GTCTGACAGGAGAACTGGAGAAGCAGATGCCAGCTGGCCGTGATCC--GCCATGTT	1356
Qy	1198	gtacgcgctgaccagcagcgcatcaagtctcaacatgaacgggcttatggcgcagccc	1257
Db	1357	GTACGAGCG--GACCAGCAGAGGATCAAGTTCATCAACATGAATGGACTCATGGAATGACCC	1415
Qy	1258	catgaaggtgtacaaagaccgcaggtcatgaacatgtggagtgagcaggagaagagac	1317
Db	1416	CATGAAGGTTCTACAGGACCGTCAGGTTACCAACATGTGGAGCAGACGAGAGGACAC	1475
Qy	1318	cttcgggagaagttcatgcagcatccccagaacctttggcctgatcgcatcttcctgga	1377
Db	1476	CTTCGCTGAGAAGTTTATGCAGCACCCCTAAGAACTTTGGCCTGATTGCTCATTTCTTGA	1535
Qy	1378	gagaagacagtgctgagtgcgctctattactacctactaagaagaatgagaacta	1437
Db	1536	GAGAAACACGGTCCGTGAGTGTGCTCTATTACTACCTGACCAAGAAGATGAATAATTA	1595
Qy	1438	taagagcctggtgagcggagctatcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1497
Db	1596	CAAGAGCTTGTTGAGCGGAGCTATCGCGCGCTGGCAAGAGC-----CAGCA	1643
Qy	1498	gcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	1557
Db	1644	GCAGCAGCAGCAGCAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	1703
Qy	1558	gaaagatggaagagaagaaagagcgcgagagagagagagagagagagagagcgagtgga	1617
Db	1704	GAAGGAGGAGAAGGAG	1763
Qy	1618	gaacgaagaagaagcactcctcaagagagaagacagacacacactcagggggagcaacga	1677
Db	1764	GAACGAGAGAGAGAACTCAGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1823
Qy	1678	cgagaagagctgtgcttccaaaggccgcaaaactcccaacagcagcaggaagacacga	1737
Db	1824	TGAGAAAGAGCGCTGCCCTCCAAAGGCCGCAAACTGCCAAACAGCAAGGCCGCCGCA	1883
Qy	1738	aggcgcatacccgctcaatggctaatgagcccaacagcagcagcagcagcagcagcagcagc	1797
Db	1884	AGGCCGTATCACGCGTCCATGSCCAACGAGGCCAACCATGAGGAGACAGCCACCCACCA	1943
Qy	1798	gcagagcgcgaagctggcctccatggagcgtgaatgagagttctcgttgacagaagaaga	1857
Db	1944	GCAAAGTTCAGAGCTGGCTTCCATGGAGATGAACGAGAGTTCCTCGCTGGACTGAGGAAGA	2003
Qy	1858	aatggaacagcgaagaaggtctcctggaacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1917
Db	2004	GATGGAGACAGCAAGAAAGGCCCTCTTGGAAATCTGGGAGGAACGTGTCGCCATTGCCCC	2063
Qy	1918	gatgtgggctccaaagctgtgtcgcagtgtaagaacttctacttcaactacaagaagag	1977
Db	2064	CATGGTGGCTCCAAAGACCGTGTCCAGTGTGAAGAACTTCTACTTCAACTACAAGAAGAG	2123

Db 4129 GTGGAGGCGCAGGAGGACTACTTACGGCGGAGGCCAAGCTCTTGAAGCAGAGGGACA 4188
Qy 4151 cctcgcgcacccacccctcagcgagcctgaccgagcgctacaagacgagcc----- 4204
Db 4189 CCACCTCCCCACGACACCTCGGGACCTGACTGAGACCTACAAGCCCGCGCCCTGGAC 4248
Qy 4205 ---ctgggccccctgaagctgaagcggcccccatgagcgccctgggtggccacggtgaagag 4261
Db 4249 COTCTGGTCCCTCGAAGCTGAAGCCGACTCACGAGGGTGTGGTAGCAACTGTGAAGGAG 4308
Qy 4262 gggggcgctccatccatgaatcccgcgagagctgcggcacacgcccgcgagctgcc 4321
Db 4309 GGGGCCCTCTATCCATGAGATGAGAGAGAGAGCTGCGCGCAGACCTGAGTATCCC 4368
Qy 4322 ctggccccgcgcctcaagagggtccatcacgagggcaccgcgctcaagtaagac 4381
Db 4369 CTGGCACCACGCGCTCTGAAGGAGGTTCCATCACCCAGGGCACCCACTCAAGTACGAC 4428
Qy 4382 accggcgctccacacactggctccaaaaagcacgagctacgctccctcatcggcagcccc 4441
Db 4429 TCTGGGGCACCCTCCACTGGCACCAAGAACACGACGTCGCGTCCATCATCGGCAGCCCC 4488
Qy 4442 ggcgagacttcccacccgctcacccgctgcatgtgatggcgagcccgcggcactgaa 4501
Db 4489 GCGCGGCTTTCCCTGCGCTGCACCGCTGGACATAATGGCTGAGCGCGCGGCACTGGAG 4548
Qy 4502 cgtgctgtacgagagagcctgaagagcgcgccagggacgcgacgagctcggggggc 4561
Db 4549 CGTGCTGCTATGAGAGAGTCTGAAGAGCGGTCAGGGACCAAGCTGGTGCGAGGGGC 4608
Qy 4562 tccattgcgggcgccccggtcattgtgctgagctgggtgaagccgcggaagagcccc 4621
Db 4609 TCCATCACACCTGGGGCTCCAGTCTGCTGCTGAATGGGCAAGCCACGCGCAAGGCCCA 4668
Qy 4622 ctgacctatgagaccacggggcaccctttgcggcaacctccacagaggttcgcccgtg 4681
Db 4669 CTGACTTACGAAGACACGGGGACCCCTTCACCACTGACCTGCCAGTGGCTGCCCTGTG 4728
Qy 4682 accatcgggagccacgcgcgctcagagggcagcctttcgtccagcaagcatcc 4741
Db 4729 ACCAGGAGGCCACGCCACGCTTCAGGAGCGACCTCCATCCAGCAAGCGCTCC 4788
Qy 4742 caggaccgaagctgaagctgacgcctcgtgagatcgccaagtcctcccgacagcaccgtg 4801
Db 4789 CAGGACCGGAAGCTGACATCTACACCCCGGAGATCGCAAGTCCCCACACAGCACTGTG 4848
Qy 4802 cccgagacacccacacccatctcgccctatgagacactgcttcggggcgtagtgac 4861
Db 4849 CCGGAGCACCCCTCACCCCATCTCCCTTCATGAGCACTTGTCTCCGGGGCGTGACTGGT 4908
Qy 4862 gtggacctgtatcgagccacatccccctggccttcgaccccaacctccatcccccggc 4921
Db 4909 GTGGACCTGTACCGTGGTTCATCCCATTTGGCTTGTACCCCACTCCATACCCCGAGGG 4968
Qy 4922 atccctctgga---cgagccgctgctactacctgccccgacacactggcccccaacccc 4978
Db 4969 ATCCCTCTGGGAAGCAGCAGCGCAGCTACTACCTGCCCCGCGCACTTGGCCCCAGCCCC 5028
Qy 4979 acctaccgcacctgaaccaacctacatcctcgcggtacctcccgacacggcgcgctg 5038
Db 5029 ACCTACCCACACCTGTATCCCACTTACCTCATTCGCGGCTACCTGACGCGCGGCCCTG 5088
Qy 5039 gagaacggcgagacatcatcaatgaactacatcaacctcgagcagatgacacaaacag 5098
Db 5089 GAGAACCCAGACCATCATCAATGACTATACCTTCGAGCAGATGCACCAACACGCT 5148
Qy 5099 gccaccgcca tggccccgagctgatatgctgaggggctctcgcggcgagctcctcg 5158
Db 5149 GCTTCGGCATGGGCCAGCGTCTGACATGCTGAGGGGTCTGTACCGCGAGAGTCTCTCG 5208
Qy 5159 ctggcactcaactacgctcggggtcccccagggatcatcgacctgtcccaagtgccacac 5218
Db 5209 CTGGCCCTCAATTATTCGGCTGGGCCCAAGAGGGCATATATCGACCTGTCCCAAGTGCCACAC 5268

Qy 5219 ctgacctgtgctgctccccgacacagggcaccagccacccagcccatggaccgcttgc 5278
Db 5269 CTGCCCGTGTGGTGCACCAACGCGAGCACCCCTGCCACCGCATCGACCGCTTGC 5328
Qy 5279 tactccccacgcgcgcacagcccttcagcagccgcccacagagctccccactctccca 5338
Db 5329 TACTTCCCACCTTGCGCCCCACCTTTCAGCAGCGCCACAGTAGCTACCGGTGTCCCA 5388
Qy 5339 ggaagtccaacacacttgacaaaacacacacacacgctcctcgcgagcgagagagac 5398
Db 5389 GGAGCCCCACTTCACTAGCTTAACCAACTGCCACATCTTTCATCGGAGGGGAACG---- 5444
Qy 5399 cgggatcagagcgagccgagcggatcgagcgggaaaaagtccatcctcagctccaccag 5458
Db 5445 -----GGAACGTGAGCGGAACGAGACAAGTCCATCTCTCAGCTTACCCT 5490
Qy 5459 acggtggagacagcaccatcttgagacactggtacagagcagagcagcgagcggcg 5518
Db 5491 ACAGTGGAGCATGCACTTGGAGACTGTGTAGAGCAGAGAGCAGAGCGGGCTGGGGC 5550
Qy 5519 agcagcgcggggggtgggggagcagcgcgcgcctccactcccatgccaccag 5578
Db 5551 AGACCGG-----CCCCGCTCCACACCCACCAG 5580
Qy 5579 cactgcacctctccctcggaccagcagatgcctccagcagagaccagtggtctcac 5638
Db 5581 CACTCGCCCATCTCCCCCGGACCCAGGAGCGCTTGCAGCAGAGGCCAGTGTGTGCAC 5640
Qy 5639 aacacagcaggaaggggtatcatcaccgctgtggagccagcaagcccacgctcctgag 5698
Db 5641 AACAGCAGCATGAAGGGGTGTGTACCTCCGTGGAAACCGGCACGCCACGGTCTGTAGG 5700
Qy 5699 tccactccactcctcactccgctgcagcagctgcacattccactccactgcacccactgc 5758
Db 5701 TCCACTCTCACCTTTCGCTGTCCGCCAGCTGCCACATTCCTCCACTGCCACCCACTGC 5760
Qy 5759 ccaactggcgggcaccctcgatgggtctaccctaccctcagcagccgcttctgtgcc 5818
Db 5761 CCACTTGTGGCACCCTTGAAGGGGTCTACCTACCTACCTCATGGAGCCGCTCTGTATCCC 5820
Qy 5819 aaggaggcccccggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgccttc 5878
Db 5821 AAGGAGACTCTCGGGTGCGCCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 5880
Qy 5879 ctgcgcaagccccagcgcgctcggggtgagcgcgctcctcccccagagagcctgcg 5938
Db 5881 CTCACCAACCCCCCGCGCG-----GGAGCGCGCTCTCTACCCAGCAAGAGCTCC 5931
Qy 5939 gagccccggccctagtgcctctgtctgtgcacgcgcacccatcgccgcacccctgcg 5998
Db 5932 GAGCCCCGATCCCTAGCACCCCGCGAGCTCCAGCCACACAGCCATCGCCGCGACCCAGCA 5991
Qy 5999 aagaacctcgacactcaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcctgc 6058
Db 5992 AAGAGCTTGGACCCACCATGCGAGTCGGAGCGCGCGCGCGCGCGCGCGCGCGCTCA 6048
Qy 6059 gaccgcaaggggaaagactcaagtaaaccttttccatccacaggaactggaactcgt 6118
Db 6049 GATTCGACCCAGGAAAGACTCAAAAGTAACCTTTTCATCCAGGAATTTGAACCTCGT 6108
Qy 6119 tctctgggtttaccac---ggcagcagctacaccccggaagggtgagccgctcagccct 6175
Db 6109 TCTCTGGGTACCACAGTGGAGCTGGCTTACAGCCCGCGATGGGTGGAGCCCATCAGCCG 6168
Qy 6176 gtgagctcaccagctctgacccaagacaaggggtctccccaagcactggaagagctcag 6235
Db 6169 GTGAGTCTCCCGCTGACCCACGACAAAGGGGTCTCCAAACCTCTCGAAAGAGTAGAG 6228
Qy 6236 aagagcaacctggagggagctcgcccaagcagcagggcccccgtaagcttgaggcg 6295
Db 6229 AAGAGCCACTTGGAAAGGGAGCTGCGGCACAGCAGCGAGGCCCATTAAGCTCAGCGCG 6288

QY 661 gtccgaagaccgcagcctgtggtgagatcatctacagcagagaaacccggaagggctgaagc 720
Db 1285 ATCAAGACCAAGCCGCTGGTTCAGATCATCTACGATGAGAACCCGGAAGCCGAAGC 1344
QY 721 tgcacatcgattctggaagcctgggcccaggtgagctgagctgtaacaaaccagcc 780
Db 1345 CCACACCGGATTCCTAGAAAGCCTGGGGCCCGAGGTGAGCTGCCTCTGTACAAACGCC 1404
QY 781 ctccgaacccggcagttatcatgagaacatcaaaataaaccagcgatgcggaagaagct 840
Db 1405 GTCCTGACACGCCAGTACCATGAANAACATCAANAATAAACAGGCGATGCGGAAGAAGCT 1464
QY 841 aatctgttactcaagagagagaaatcaagctcgggaacaaatgggaagcagaagtctgcca 900
Db 1465 GATCTTGTACTTAAAGCGGAGGAACACCGCGCAAGCAGTGGGAACACAGCGCTTCTGCCA 1524
QY 901 ggcctatgacacagctcatggagcccttggaaaaaaggtgagcgcatcgaaaaaacccc 960
Db 1525 GCGCTATGACCAAGCTCATGGAGGCGCTGGGAGAAAGGTAGAGCGCATAGAGAAACAATCC 1584
QY 961 ggcgcgcggcgccaaagagagcaaggtgcgcgagttactacgaaagcagttccctgagat 1020
Db 1585 GCGAAGGAGGCGCAAGGAGAGCAAGGTGAGGAGGTACTACGAGAAACAGTTCGCCGAGAT 1644
QY 1021 ccgcaagcagcgagcgtgcaggagcgcatgcagagcaggggtggccagcgggcgagttgg 1080
Db 1645 CCGCAAGCAGCGGAGCTGCAGAGCGCATGCAGAGCAGGCTGGCCACAGCGTGGCAGTGG 1704
QY 1081 gctgtccatgtgcctc 1140
Db 1705 GCTCTCCATGTGGGTGCCCGCAGTGAGCATGAGTTTCTGAGATCATGTGTTGCTGTC 1764
QY 1141 agacgagagaaacctggagaagcagatgcgcagctgcccgtgacccgcacactgctata 1200
Db 1765 TGAGCAGAGAAACCTTGAGAGAACAGATCGCCAGCTGCCGTGATCCGCCCATGTTGTA 1824
QY 1201 cgaagctaccagcagcgcatcaagttcatcaacatgaacgggcttatggccgaaccccat 1260
Db 1825 CGACCGGACCAAGCAGAGAGTCAAGTTTCAATCAACATGAATGGACTCATGGATGACCCCAT 1884
QY 1261 gaaggtgtacaaagaccgcaggtcatgaacatgtgagtgagcagaggaagagacatt 1320
Db 1885 GAAAGTCTACAAAGGACCTCAGGTTTACCAATGTGGAGCAGCAGAGAGGAGACACCTT 1944
QY 1321 ccgggagaagttcatgcagcaccgaactttggcctgcatcgatcatcttcctggagag 1380
Db 1945 CCGTGAGAAGTTATGACGACACCTTAAGACTTTGGCCCTGATGGCTCATTTCTCGGAGAG 2004
QY 1381 gaagacagtggtgagtcgtct 1440
Db 2005 AAAGACGGTCGTGAGTGTCTCTATTACTTACTGACCAAGAAAGTAATAAATTACAA 2064
QY 1441 gagcctgtgagcagagactatcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1500
Db 2065 GAGCTTGTGAGGGGAGCTATCGCGCGCTGGCAAGAGCCAGCAGCAGCAGCAGCAGCA 2124
QY 1501 gcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1560
Db 2125 ACAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2172
QY 1561 agatgagaag 1620
Db 2173 GGAGGAGAAGGAGAAGGAGAAGGAGGCGGCACAAAGAGAGAGAGAGAGAGAGAGAGAG 2232
QY 1621 cgaagaagaacctctcaag 1680
Db 2233 CGAGAGGAAGAACTACAGAGGAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2292
QY 1681 gaag 1740
Db 2293 GAAAGAGGCCCTGCCATCCAAAGGCCGCAAACTGCCAAACAGCAAGGCCCGCCCAAGG 2352
QY 1741 ccgcataccccgctcaatggttaatgaggcccaacagcgagagagagagagagagagag 1800

Db 2353 CCGTATCATCGCGCTCCATGGCCACAGAGCCAAACCATGAGGAGACAGCCACCCACAGCA 2412
QY 1801 gaggcgcagctgctccatccatgagcgtgaatgagagttctcgtgagcagagaagaataa 1860
Db 2413 AAGTTTCAGAGTGGCTTCATGGAGATGAAGCAGAGTTCCTCGCTGGACTGAGGAAGAT 2472
QY 1861 ggaacacgcaagaaaggtctcctggaacagcgccgcaactggtcgccatcgcccgat 1920
Db 2473 GGAGNACAAAGAAGGCCCTCTGGAACATGGGAGGAACCTGGTCAGCCATTCGCCGAT 2532
QY 1921 ggtgggtccaagactgtgcgagctgaagaactctacttcaactacaagaagagca 1980
Db 2533 GGTGGCTCCAAGACCGTGTCCAGTGTGAAGACTTCTACTTCAACTACAAGAAGAGCA 2592
QY 1981 gaacctcgatgagatctctgagcagcacaagctgaagatggagaagagagagagagcgcg 2040
Db 2593 GAACTTGACCAAAATCTTTCAGCAGCACAAAGTAAAGATGGAGAGAGAGAGAGAGCTCG 2652
QY 2041 gagaaagaagaagc 2100
Db 2653 GAGGAAGAAGAAGAGACCCACAGCTGCGCGAGCGAGGAGAGACAGCTTCCACCTGCCGC 2712
QY 2101 ggagatgagagatggaggcgtcggtgagcggaatgagggagagatggtggagga 2160
Db 2713 TGAGGACCAAGAGATGGAAGCATCAGGCGCAAGTGCCAAATGAGGAAGAGCTGGCGAGGA 2772
QY 2161 ggcgtgaacctatactgctctggaatgaggtgcccaga---ggggaaatcagtgagcc 2217
Db 2773 GCGAAGACCTCACAGGCTCTGGGAATGAGGTTCCAGAGTTGGGAGTGAGTGGGCC 2832
QY 2218 agcactgtcaacaacagctcagacacagagatccccctctctcctacactgagagccg 2277
Db 2833 AGCTGCTGTCAACAAACAGCTCTGATCTGAGAGTGTCCATCCCGCGCTTCAGAAAGCCAT 2892
QY 2278 caaggacacagggcagaatgggccaagccccagccacccctggcgcgcgagggccacc 2337
Db 2893 GAAGGACAC-----TGGGCTTAAACCCACTGGCACTGAAGCATTTGCCGCTGCCAC 2943
QY 2338 cccagggcccccaccccccacacagagagagacatccccgggcccccatgagccacccccgc 2397
Db 2944 CCAGCCACCTTCTCTCTCCAGAAAGAACCGGAGTAGTAGCCCTGCTGAGCCCTCCCCAGT 3003
QY 2398 ctctgaagcccccagagccccctagccccccacagcccccatcgccctctgacactcc 2457
Db 3004 CCTGTATGCCAGTGG---CCACCATCCCCAGAGCCTTCCCATCACTGCGGCAACCCCC 3059
QY 2458 tcctgtgtcccccaagggag 2517
Db 3060 GCCTACTGTGCACAAGGATGAACAAGAGCCCGCGCTGCTCCAGCTCCCCAGACAGAGGA 3119
QY 2518 gggggag 2558
Db 3120 TGCCAAAGAGAGCAGAGTGTAGGCGCGAGGAGATCATGTGGGAAAGCCAGAGAGAGCCG 3179
QY 2559 tggacacaggaagcgag 2618
Db 3180 AGGCTCTGAGAGAGCCCCCGGAGAGTGTAAAGAGTGACCAACAAGGAGGAGACCGAGGAG 3239
QY 2619 gggcggc---caagggcaagggagcgagcgctgagggccacgcgcgcgcgcgcgcgcgcgc 2675
Db 3240 AGCCTGAAGACAAAAGCCAAAGGCCACAGAGGCCATTGAACCTGTGTGTGAGGACACACTTA 3299
QY 2676 aggcag 2735
Db 3300 AG---GTGGAGAGGCTGTAGTAGAAGCAGCTGTGACCAAGGTTTCCAGGTCAGGTGCCA 3356
QY 2736 cccaggacagcagctccagtgactctacctgagtgagagagagagagagagagagagagag 2795
Db 3357 CCCAGGACAGTACTTCAGTGCCACTTCAGTGCGAGTGAGTGAGTGAGGACCAACCCAGAGG 3416
QY 2796 gcgacaagaacccgctgctgttccccagggccagcctctctacccccactggtggagcccc 2855

Db 3417 CTGACAGGGAGGCTGCTGTACCAAGGCCAGCCTCTCTACCCCGGTGGAGATCCCC 3476
Qy 2856 gggcaatgctcaccacagcaactggaccctgaacagctgaagcagagcggtg 2915
Db 3477 GGGCCAGTACCTCGCCCCAGAAAGCGCTGGACCTGAAGCAGCTGAAGCAGGAGCAGCG 3536
Qy 2916 ccatccccccatcaggtaccacaaagtccatgagccccccgggagcagctccca 2975
Db 3537 CCATCCCCCTATCCAGTCCACCAAGGTCATGAGCCCCCGGGAGGACACAGTACCCC 3596
Qy 2976 ccaagcaagctccccacgccccacacgcgcaaaactgcagcggagcgagcgccc 3035
Db 3597 CAAGCCAGTTCCCTCTGCTCCACCCAGCAGCAGCTACAGCCAGAGGGTGAGTGT 3656
Qy 3036 ctacagcagctggcagcagccccggggcgaagagcagcagcccccccgccagaca 3095
Db 3657 CTCAGCAGTCGGGAGGAAGTCACTGTGCAAGTCCCGAGGCCAGTGTCTCTGCCGAGA 3716
Qy 3096 agga-----ggccttgcagccgagggcccaagagctgctg 3131
Db 3717 AAGAGGCAGAGAAACCCGCATTCTTTTCGGGCTTTCCCAACTGAGGGCCCAAAGCTACCGA 3776
Qy 3132 gggaccccccttgctgagcttcggcctgccccttcccctgcccccccgtaggtatca 3191
Db 3777 CTGAGCCCCACAGCTGTGTATCGGGCTTGCCCTTCCCATCCCTCCACGGGAGGTGATCA 3836
Qy 3192 aggcctcccccatgcgccccgaacccctcagcctctctcagctccacctgtaacccac 3251
Db 3837 AGACTTCCCAACAGCCGCTGACCCCTTGCCCTTCTCTACACACCCCGGTACCCGCG 3896
Qy 3252 tgccttgggctcccatgacactgccccggcctgctctccgccccccacccactctcca 3311
Db 3897 TGCCCTCGGGCTCCACGATAGTGCAGGCCGCGCTCTGCCAG---TCCCCCATCTCTA 3953
Qy 3312 acccgctccccctatctctctgccaagacccccagcgctcctcgagaggcaaatagtg 3371
Db 3954 ACCCCCAACCCCTCATCTCCCTGCAAGATCCCGGCGTACTTGAGAGGAGCTGGGTG 4013
Qy 3372 ccatctcc---caaggaatgctggtccagctccagctccctactcagagcagcgaagg 3428
Db 4014 CCATCTCCACAGCAGGGATGTCACTCCAGCTTCGTGTCCCTCACTCAGAGCATGCCAAG 4073
Qy 3429 cccgggtggccctgtaccatggggtgccccctgccccatggaacccccaaaagtggcac 3488
Db 4074 CCCCATGGGCCCTCTACCATGGGCTTGCCCTTGCCGTGGACCTAAGAAGCTG---- 4129
Qy 3489 ccttcagcgagtgaaagcagagcagctgtccccacggggccaggctgggcccacggaga 3548
Db 4130 ----- 4129
Qy 3549 gcttgggggtgcccacagccagaggcgctcgtgtagagggacagctctgggctcag 3608
Db 4130 -----GGACACACTGGGCTCCG 4148
Qy 3609 ttccggggcgaagcataccaaaagcattcccagcacacgggtgcccctcgacagcgcca 3668
Db 4149 CCACCAGTGGGAAGCATCACCAGGGCTTCCCGAGTACCCGG-----CTCAGACAGGCC 4202
Qy 3669 tcacataccggcgttccatacccaacgcgacgcagctgacgtctctgtacaaaggacaca 3728
Db 4203 CCAGCTCAGAGGCTCTATACCCACGCACGCCCGCCGACAGAGTCTCTTACAAAGGTATCA 4262
Qy 3729 tcaccagatcatcggcagagacagcccgagctgcttgacccggcgccggagagacagcc 3788
Db 4263 TCAGCAGGATCTGTGGTGAGGACAGCCCAAGTCGCTTTGACCGGGCAGAGGACACCC 4322
Qy 3789 tggccaaggccatcctacgaaggaaggaagggccacgcttcttcttatagggtg 3848
Db 4323 TGCCCAAGGGCCATGTATCTATGAGGGCAGAAAGGCCAGCTCTATCTTATGAGGTG 4382
Qy 3849 gcatgtctgacccagctgtcccaaggagcagcggaagcagctcaggaaccccccatg 3908
Db 4383 GTATGTCCGTGTACAGTGTCTAAGGAGGATGGAAGGAGCAGCTCGGGGCCACCCCATG 4442

Qy 3909 agagcgccgcccccaagcgacacctatgacatgatgagggcgcgctgggcaagcactc 3968
Db 4443 AGACTGCGCGCCCTTAACGCGACCTATGACTATGATGAGGGCGCTGTAGCAGGACTGTCA 4502
Qy 3969 cctcagcagcatcgaaggtctcatgggctgcatcccgccgagcgacacagcccc 4028
Db 4503 CCTCAGCCAGCATAGAGGACTCATGGCGCGGCCATCCTTGAGCAGCA-----CAGCC 4556
Qy 4029 accactcaagagcagcaccacatcccggttccatcacacaagggatccccgttct 4088
Db 4557 CCCACTCAAGSAGCAGATCACATCCGAGGCTCCATCAGCAAGGCATCCCGAGGTCT 4616
Qy 4089 acgtgagcgacagagactacctcgtcgaggcccaagctcctaaagcgaggga 4148
Db 4617 ATGTGAGCGCAGAGGACTACTTACGGCGGAGGCCAAGCTCTTGAAGCAGAGGGA 4676
Qy 4149 cgcctccgccccccacgcgcctcacgggacctgacgagggcctacaagacgaagccc--- 4204
Db 4677 CACCACCACCCACACACACCTCGGAGCTGACTGAGACCTTACAAGCCCGCCCTGG 4736
Qy 4205 -----ctgggccccctgaagctgaagccggccccatgagggcctggtggccacgggtgaagg 4259
Db 4737 ACCCTCTGGGTCCCTGAAGCTGAAGCGGACTTCAAGAGGGTGTGTAGCACTGTGAAG 4796
Qy 4260 agcgggcgctccatccatgacatcccgcgagagctcgcgacacgcccagctgc 4319
Db 4797 AGCGGGCGCTCTATCCATGAGTCCCGAGAGAGGAGTGGCGCGCACACCTGAGCTAC 4856
Qy 4320 ccttggcccccgggcgctcaaggagggtcctcatcacgagggacccccgcctcaagtacg 4379
Db 4857 CCTTGGCACACCGGCTCTGAAGAGGGTTCATCACCCAGGGCACCCACATCAAGTAGC 4916
Qy 4380 acacggcgcttccacacacggctcccaaaagcagacgtaagctcctccatcgagccc 4439
Db 4917 ACTCTGGGGCACCTTCCACTGGCACCAAAACACAGAGTGGCTCCATCATCTCGCAGCC 4976
Qy 4440 cggcgcgagctgtcccaacccgtgcacccgctggtatgtagcgccgacccggcactgg 4499
Db 4977 CCGCGCGGCTTTCCTTGCCTGCACCCGCTGGACATATATGGCTGACGCCCGGCGACATG 5036
Qy 4500 aacgtgcctgtacagagagcgtgaagcgccggccgagcccgagctcggggg 4559
Db 5037 AGCTGCTGCTATGAAGAGTCTTGAAGAGCGCTGACGGACACAGCATGTGTGACGGG 5096
Qy 4560 gctccattgcgcgccgcccccgctcattgtcctgagctgggtaagcgcgagagcc 4619
Db 5097 GCTCCATCACAGTGGGCTCCAGTCTGCTGCTTGAACCTGGGCAAGCCACGGCAAGGCC 5156
Qy 4620 ccttgacctatgagcacacggcgccacctttgcccggccacctcccacgaggttcgccc 4679
Db 5157 CACTGACTTACGAAGACACCGGGCACCTTACACAGTCACTGCGCAGCTGCTCCCTG 5216
Qy 4680 tgacctgaggagccccacgcgcctgcagagggcgagccttctgtccagcaagcat 4739
Db 5217 TGACCAGAGGAGGCCACGCCCTTCAAGAGGAGCGCTCTATCCAGCAAGGCT 5276
Qy 4740 cccagacccgaagctgagcgtgacgctgtgagatgcgaagtccccgcacacagccc 4799
Db 5277 CCCAGGACCGAAGTGCATCTACACCCCGGAGATGCCAAGTCCCGCACACAGCATG 5336
Qy 4800 tggcgagcacacccacaccccatctgcctatgacacctgttcggggcggtgagtg 4859
Db 5337 TGCCGACACACACCTTACCCCATCTCCCGTATGAGCATTTGCTCGGGGCGGTGACTG 5396
Qy 4860 gcttggacctgtatgcgagccacatccccctggccttcgacccccctccatcccccg 4919
Db 5397 GTGTGAGCTGTACCGTGTGTACATCCCATTTGGCCTTTGACCCCACTTCCATACCCGAG 5456
Qy 4920 gcatcctcttga-----cgagccgctgctactactgccccgcacacactggcccccaacc 4976
Db 5457 GGATCCCTCTGGAAGCAGCAGCGCGAGCTTACTACTTGCCTCGCGGCACTTGGCCCCCAGCC 5516

QY 4977 ccactaccgcacactgtaccacacactacactccgcggtaccccgagacacgagcggcg 5036
Db 5517 CCACCTACCCACACCTGTACCCACCTTACCTCATCCGCGGTACCTGCACAGCGGGGCC 5576
QY 5037 tggagaacgagcagacatcatcaatgactacatcaactcgcagcagatgacacaca 5096
Db 5577 TGGAGAACCGCAGACCATCATCAATGATACTATCATCACCTCGCAGCAGATGACCAACAACG 5636
QY 5097 cggccaccgcatgcccagcagactgatatgctgaggggctctcgcgcgagctct 5156
Db 5637 CTGCTCCGCCATGCCCCAGCGTGTGACATGCTGAGGGGTCTGTCACCGGAGAGTCCCT 5696
QY 5157 cgtggaactcaactacgtcgtgggtcccgagggcatcatcgacctgtcccaagtgcac 5216
Db 5697 CGCTGCCCTCAATTTATGCGCTGSCCCCAAGAGGATTTATCGACCTGTCCCAAGTGCCAC 5756
QY 5217 acctgctgtgctgtgcccgcgacacacagcagcaccaccccgacacgcgccttg 5276
Db 5757 ACCTGCCGTGTGTTGTGACCAACAGCCAGGACACCCCTGCCACCGCCATCGACCGCTTG 5816
QY 5277 cctacctccacgcgcccagcgccttcagcagccagccacagactccccactctccc 5336
Db 5817 CCTACTCCCCACTGCGGCCCCACCTTCAGCAGCGGCCACAGTAGCTACCGCTGTCCC 5876
QY 5337 caggaggtccacacacttgacaaaacacacacacacacacacacacacacacac 5396
Db 5877 CAGGAGGCCACCTACCTAGTTAAACCACTGCCACATCTTCATCGGAGCGG----- 5929
QY 5397 accggggtcagagcgggagcggggtcgggagcggggaagaaagtccatctcagtcacac 5456
Db 5930 -----GAACGGGAACGTGAGCGGGAACGAGAACAGTCCATCTCATCGTCTACCA 5978
QY 5457 cgacggtgagcagcaccatctgagagactggtacagagcagagcagcgagcagcg 5516
Db 5979 CTACAGTGGAGCATGCACCCATCTGGAGACCTGTGTACGAGCAGAGCAGCGGGGTGGG 6038
QY 5517 gcagcagcggcggggtggggagcagcagcgcgcgcgcgcctccactcccatgcccacc 5576
Db 6039 GCAGCAGCGGC-----CGGCTTCCACAGACCCAC 6068
QY 5577 agcactgcacatctccctcggaccagagatgcctccagcagagacacagtgcttc 5636
Db 6069 AGCATTGCCCATCTTCCCCCGGACCCAGGACGCTTTCGACAGAGGCCCCAGTGTGCTGC 6128
QY 5637 acaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5695
Db 6129 ACAACACGAGCATGAAGGCGTGTGTCACCTCGTGGAAACCGGCACGCCAGGCTCTGA 6188
QY 5696 -----aggtcaactccactcctcactccgcgttcgcccagctgcacattccacactg 5747
Db 6189 GGTGGGCCAGGTCCACCTTCCACCTTTCGCTGTCCGCTTCCGCTTCCACATTCACACCTG 6248
QY 5748 ccaccactgcacactgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5807
Db 6249 CCACCCATGCTGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 6308
QY 5808 tcttgctcccaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 5867
Db 6309 TCTCTTTACCAAGAGACCTCTCGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6368
QY 5868 gccatgcttctcgcgaagccccagccccgcctcgcggtgagccccgcctcctccccc 5927
Db 6369 GCCATGCTTTTACCAAAACCCCGGCGG-----GGAGCGCCGCTTCTTCCACCCA 6419
QY 5928 gcaagggctcggagccccgcctcagctcctgctcctgctcctgctcctgctcctcctcct 5987
Db 6420 GCAAGAGTTCGAGCCCGATCTCTAGCACCCCGCCAGCTCCAGCCACACAGCCATCGGCC 6479
QY 5988 gcaacccctgcgaacactcgcacatccacacagccccgcgagccccgcgagccccgcg 6047
Db 6480 GCACCCAGCAAAACCTTGACCCCAACCTTGCACCCCAACCTTGCAGTCCGCGCGCGCGCG 6536
QY 6048 cctcggcctcggacccgcagcgggaaagactcaagtaaaccttttccatccaggaac 6107

Db 6537 CCTCGGCTCAGATCTGCACCGAGAAAGACTCAAGATAAACCCTTTTCCATCCAGGAAT 6596
QY 6108 tggaaactcgttctctggtttaccac-----ggcagcagctacagccccgaaaggggtggagc 6164
Db 6597 TGGAACTCCGTCTCTGCTGGGTTTACCACAGTGGAGCTGGCTACAGCCCCGATGGGTGGAGC 6656
QY 6165 cgtcagccctgtgagctcaccagctgacccagcagcagcagcagcagcagcagcagcagc 6224
Db 6657 CCATCAGCCCGGTGAGCTTCCCCAGCCTTGACCAGCAAGGGGCTCTCCAACATCTCTGG 6716
QY 6225 aagagctgacaaagcagccactgaggggagctgagcagcagcagcagcagcagcagcagc 6284
Db 6717 AAGAGCTAGAAAGAGCCACTTGGAAAGGGAGCTGCGCACAAGCAGCAGCGCCCATGA 6776
QY 6285 agcttgccggggagcgcgcacactccacactcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6344
Db 6777 AGCTCAGCGCGGAGGCTGCCATCTCCACATCTCGGCGCACTGCGCGAGAGCCAGCCCT 6836
QY 6345 cgtccagccgctcctcagaccgccaggggtcaagagctcaccagcgggtggttcaccc 6404
Db 6837 CATCAGCCCACTCTCTCAGACTGCCCCAGGATYCAAAAGTTCACAGAGGTTGTTCAACC 6896
QY 6405 tggccagcacatcagtgaggtcatcacagggactcacccgcgcgcgcgcgcgcgcgcgcgc 6464
Db 6897 TGGCTCAGCATCAGCGAGGTCTATTACGACGACTACACGCGCCACCCCGCAGCAGC 6956
QY 6465 tcagcgaacccctgc 6524
Db 6957 TCAGTGGCGCTTCCCCGCCCTCTCTACTCTTTCGCCGAGCCAGCTGCGCTGTCTGG 7016
QY 6525 acctccgc 6584
Db 7017 ATCTTCGCGCGCCACCCAGTGAACCTCTACCTTCCACCCCGCCAGCCATGCGACCCCGCC 7076
QY 6585 gtgctcccccacagcgaaagggggaaggtctccagagcgaacaaagcgtcggtct 6644
Db 7077 GGGATCCCCCACAGTGAAGGGGGCAAAAGTCCCCAGAACCCAGCAAAACATCGGTCC 7136
QY 6645 tgggtggtgagcagcaggtattaaacctgtgtccacagcaggggcatgacagagcagcag 6704
Db 7137 TGGCAGCAGCGAGGATGCCATGTAGCCCTGTGTCCTCCACAGAGGGCATGATGAGCCAG 7196
QY 6705 ggcactcccgagtgctgtgtaccgcgtgctgacccggatggggaacagcagcagcagcag 6764
Db 7197 GACATGCTCGGAGCACTGCGTACCCACTGCTGTATCGAGACGGGGAACAGGCGGAGCC-- 7254
QY 6765 gcaggtggttccaaagtctccagggcaacacagcagcagcagcagcagcagcagcagcagc 6824
Db 7255 -CAGGATGGGTCTAGAGTCTCCAGGCAACACCCAGCCGCCAACCTTCTTCAAGTAAAGC 7313
QY 6825 tgcagcagcagcaactccgccaatggtcaagtccaagaagcagagatcaacaagaagctga 6884
Db 7314 TGACTGAGAGCACTCCGCCATGTGTAAGTCGAGAGCAGGAGATCAACAAGAATCA 7373
QY 6885 acacccacaacccggaatgagcctgaatacaataatgcagcagcagcagcagcagcagcagc 6944
Db 7374 ACACCCACAACCGGAACGAGCCAGAAATACAATATTGGCCGCTGGGACGGAAATCTTCA 7433
QY 6945 atatcccgcccatcacggaacagcccttatgacctatagaagcagcagcagcagcagcagc 7004
Db 7434 ACATGCCCGCCCATCCTGGAGCAGGCTTATGACCTGTAGAAAGCCAGCGGCTGCAAGAAC 7493
QY 7005 atgcagcagcaacatggggtgagcagcagcagcagcagcagcagcagcagcagcagcagc 7064
Db 7494 ACGCCAGCACCAACATGGGGCTAGAGGCCATTTATTAGAAAGCAGCTCATGGGTAATATG 7553
QY 7065 accagtggaagagtcgccgc 7124
Db 7554 ATCAGTGGAAAGAGCCCCCGCGCTCGCGCGCAATGCTTTTAAACCCCTCTGAATGCCAGCG 7613
QY 7125 ccagcctgccc- - -gctgctatgccataacgcctgtgcagcagcagcagcagcagcagcagc 7181

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Qy 6346 gtccagccgctgtccagacgccccctactcttccctgggcccagctccccct 6405
Db 612 gtccagccgctgtccagacgccccctactcttccctgggcccagctccccct 671
Qy 6406 gcccagacacatcagtgaggtatcacacagactacacccggaacccacacagct 6465
Db 672 gcccagacacatcagtgaggtatcacacagactacacccggaacccacacagct 731
Qy 6466 cagcagccctgtccgccccctactcttccctgggcccagctccccct 6525
Db 732 cagcagccctgtccgccccctactcttccctgggcccagctccccct 791
Qy 6526 cctccgcccacccagtgactctactctccgccccggaacctgtgtcccccgcgcg 6585
Db 792 cctccgcccacccagtgactctactctccgccccggaacctgtgtcccccgcgcg 851
Qy 6586 tggctccccccacagcgaagggggaaggtctccagacacaaagacgttgttt 6645
Db 852 tggctccccccacagcgaagggggaaggtctccagacacaaagacgttgttt 911
Qy 6646 gggtagtgagcaggtattgaacctgtgtaccgggtgtggaacacagcagccag 6705
Db 912 gggtagtgagcaggtattgaacctgtgtaccgggtgtggaacacagcagccag 971
Qy 6706 gcactcccgagtgctgtaccgctgtgtaccgggtgtggaacacagcagccag 6765
Db 972 gcactcccgagtgctgtaccgctgtgtaccgggtgtggaacacagcagccag 1031
Qy 6766 caggtatgggtcccaagttccaggaacacacagcagccagcctttccagcagct 6825
Db 1032 caggtatgggtcccaagttccaggaacacacagcagccagcctttccagcagct 1091
Qy 6826 gaccagagcaactccgcatgtgaatccaaagacgaagatcaacaaagacgttaa 6885
Db 1092 gaccagagcaactccgcatgtgaatccaaagacgaagatcaacaaagacgttaa 1151
Qy 6886 caccacacccggaatgagctgaatacaatacagcagcctgggagcgagatctcaa 6945
Db 1152 caccacacccggaatgagctgaatacaatacagcagcctgggagcgagatctcaa 1211
Qy 6946 tatgcccgcataccggaacagccttatgacctatagaagccagcgggtcaggaaca 7005
Db 1212 tatgcccgcataccggaacagccttatgacctatagaagccagcgggtcaggaaca 1271
Qy 7006 tccagacccaatgggctggagccataattagaagcactcatgggttaatatga 7065
Db 1272 tccagacccaatgggctggagccataattagaagcactcatgggttaatatga 1331
Qy 7066 ccagtgggaagagtcgcccgctcagcgccaatgttttaacctctgaatgccagtcg 7125
Db 1332 ccagtgggaagagtcgcccgctcagcgccaatgttttaacctctgaatgccagtcg 1391
Qy 7126 cagcctgcccgtgctatgcccataaccgctgtgacggaagcaggtgaccacactcac 7185
Db 1392 cagcctgcccgtgctatgcccataaccgctgtgacggaagcaggtgaccacactcac 1451
Qy 7186 ctccgaggtggcggaagccaaaggtctctggaagccagcagcgaagaaagccaa 7245
Db 1452 ctccgaggtggcggaagccaaaggtctctggaagccagcagcgaagaaagccaa 1511
Qy 7246 gtccccggccccggcctggcatctggggaacggccacccctctgtctcctcagtgacac 7305
Db 1512 gtccccggccccggcctggcatctggggaacggccacccctctgtctcctcagtgacac 1571
Qy 7306 gaaggaagactgaacccggaagccgctcaccacacccgctgtgggagcagccctc 7365
Db 1572 gaaggaagactgaacccggaagccgctcaccacacccgctgtgggagcagccctc 1631
Qy 7366 gtccgaggttccagccatccccctacaaacccctgacatcgagcgtgagcgaggtgt 7425
Db 1632 gtccgaggttccagccatccccctacaaacccctgacatcgagcgtgagcgaggtgt 1691

Qy 7426 catggtttccccccccccccccggtcctccccgcgggagcggccccctcgtggtccccca 7485
Db 1692 catggtttccccccccccccccggtcctccccgcgggagcggccccctcgtggtccccca 1751
Qy 7486 ccaagcctggagcagagagcccaagcactcgtcgtcagtagtacagaaactctcga 7545
Db 1752 ccagcctggagcagagagcccaagcactcgtcgtcagtagtacagaaactctcga 1811
Qy 7546 cagcagtgactgaacacagcggcgggggcgggcgggcggtteuersperluetcaggtccc 7605
Db 1812 cagcagtgactgaacacagcggcgggggcgggcgggggg-----CGGTGTAGGTCCC 1860
Qy 7606 agcagccacaggaagccctcagggagcggcggtcgcgactcccccaacaaagga 7665
Db 1861 agcagccacaggaagccctcagggagcggcggtcgcgactcccccaacaaagga 1920
Qy 7666 aggaagccctcagtcgctcgcctccatccatcgttccgtcgaagccggcactccttg 7725
Db 1921 aggaagccctcagtcgctcgcctccatccatcgttccgtcgaagccggcactccttg 1980
Qy 7726 cctgtctaaagcccttaactaaagactcccccccggtcgtcgcagaccttaactca 7785
Db 1981 cctgtctaaagcccttaactaaagactcccccccggtcgtcgcagaccttaactca 2040
Qy 7786 ggggatgtttacctgggtcgtcgggaagggagggcgggcgggcgggcgagcgag 7845
Db 2041 ggggatgtttacctgggtcgtcgggaagggagggcgggcgggcgagcgagcgag 2100
Qy 7846 gctgtgtgagcagcagcagcagcggcgggcgggcgggcgggcgagcgagcgag 7905
Db 2101 gctgtgtgagcagcagcagcggcgggcgggcgggcgggcgagcgagcgagcgag 2160
Qy 7906 gcaactccagccactcctcccccgaatgcatattgaaccaaagcttaaaagactcgt 7965
Db 2161 gcaactccagccactcctcccccgaatgcatattgaaccaaagcttaaaagactcgt 2220
Qy 7966 cagcccccgccctcccccgctccccatccccgcttagcgtctggacagatgagcgca 8025
Db 2221 cagcccccgccctcccccgctccccatccccgcttagcgtctggacagatgagcgca 2280
Qy 8026 ggcctgtccagcccccgctccccgctccccgctccccgagcagcagcagcag 8085
Db 2281 ggcctgtccagcccccgctccccgctccccgctccccgagcagcagcagcag 2340
Qy 8086 atgtctgaaacaaagtcagcaggtggtggtggtggtggtggtggtggtggtggtggt 8145
Db 2341 atgtctgaaacaaagtcagcaggtggtggtggtggtggtggtggtggtggtggtggt 2400
Qy 8146 gaaaggtgctccgaggtgagctgtttttttcacacatcgttgcgagcagcagcagcag 8205
Db 2401 gaaaggtgctccgaggtgagctgtttttttcacacatcgttgcgagcagcagcagcag 2460
Qy 8206 ggaagggcagatgataatgagtggtgtttacaggggtatattttttgatacctcaatga 8265
Db 2461 ggaagggcagatgataatgagtggtgtttacaggggtatattttttgatacctcaatga 2520
Qy 8266 attaatcagatgttttacgaaagagacttaccagtagtattcgtcgtcgtcgtcgtcgt 8325
Db 2521 attaatcagatgttttacgaaagagacttaccagtagtattcgtcgtcgtcgtcgtcgt 2580
Qy 8326 gatctcgttaccgttcaagagcgtgtgagcagcagcagcagcagcagcagcagcagcag 8385
Db 2581 gatctcgttaccgttcaagagcgtgtgagcagcagcagcagcagcagcagcagcagcag 2640
Qy 8386 caggaacaaagggcgggagcgtgtgagcagcagcagcagcagcagcagcagcagcagcag 8444
Db 2641 caggaacaaagggcgggagcgtgtgagcagcagcagcagcagcagcagcagcagcagcag 2700
Qy 8445 cctgggcagagtaaatcagatcgtattcgtggtggtggtggtggtggtggtggtggtggtggt 8504
Db 2701 cctgggcagagtaaatcagatcgtattcgtggtggtggtggtggtggtggtggtggtggtggt 2760
Qy 8505 ttctgtcatttacacagcgtgttcttaataaaaaagcgaattataactcccaaaaaaa 8564

```
|||||
Db 2761 TTCTGCTATTACACACGTCGTTCTTAATATAAAAGCGAATTATATATCCAAAAA 2820
QY 8565 aaaaaa 8571
Db 2821 AAAAAA 2827

RESULT 8
S83390 2930 bp mRNA PRI 28-MAR-1997
LOCUS T3 receptor-associating cofactor-1 [human, fetal liver, mRNA, 2930 nt].
DEFINITION
ACCESSION S83390
VERSION S83390.1 GI:1911769
KEYWORDS human fetal liver.
SOURCE
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2930)
AUTHORS Sande,S. and Privalsky,M.L.
TITLE Identification of TRACs (T3 receptor-associating cofactors), a family of cofactors that associate with, and modulate the activity of, nuclear hormone receptors
Mol. Endocrinol. 10 (7), 813-825 (1996)
JOURNAL 96408715
MEDLINE
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 179975] from the original journal article.
This sequence comes from Fig. 4.
FEATURES
source
1..2930
/organism="Homo sapiens"
/db_xref="taxon:9606"
456..2765
/gene="T3 receptor-associating cofactor-1, TRAC-1"
456..2765
/gene="T3 receptor-associating cofactor-1, TRAC-1"
/note="T3 receptor-associating cofactor-1, TRAC-1"
/note="This sequence comes from Fig. 4; Authors report detecting several TRAC alternative transcripts; TRAC-1"
/codon_start=1
/product="T3 receptor-associating cofactor-1"
/protein_id="AAB50847.1"
/db_xref="GI:1911770"
/translation="MAQRADMLRGLSPRESSLALNYAAGRGIIDLSQVPHLPVLVPP
TPGTPATAMDRLAYLPTAOPFSSRHSSSPLSGPPTHLTPTTTSSRERDRDRER
DRDRERKSLTSTTVEHAPLWRPTEQSSGSGSGSGSSRSPASHAHQHP
ISRTDQALQORPSYLNHMGKLIITAVESTPTVLRSTSTSPVRPAATFPATHCP
LGTLDGVTPLMEPVLLPKPEARVARPERPRADTGHAFKAPKSPGLSPSPSKG
SEPRPLVPVYSGHTARTPAKPLAPHASPDPPAPASADPHREKTSKPFISQEL
ELRSGLVHGSYSVEGPEVPSVPSPLQTPGVKGHORVVTLAOHISVITQDYTRHP
KLGEAAHLPLRLPESOPSSPLQTPGVKGHORVVTLAOHISVITQDYTRHP
QQLSAPLAPLYSFGACSPVLDLRLPPSDLYLPPDPHAPAGSPHSEGGKSPEN
KTVSLGGEGDIEVSPPEGTEPGHSRAVILLYRDGEQTEPSRSGSPNTSQP
PAFFKLTNSAWKSKQEIINKLHNREPEYNISQPFIEFNPAITGLMT
YRQAOVHEHSTNNGLEAIIRKLMGGGKAKVSPSRKASAPAGLASGRRPSV
SVVHSEGDENRRPTLNRVWEDRPSASSTPFYPLNLRQAGVNASPPPPGLPAGS
GPLAGAHAWDEEPKPLLCQSQYETLSDSE"
BASE COUNT 586 a 1149 c 780 g 415 t
ORIGIN

Query Match 30.6%; Score 2618.6; DB 97; Length 2930;
Best Local Similarity 94.6%; Pred. No. 0;
Matches 2786; Conservative 3; Mismatches 9; Indels 146; Gaps 2;

QY 4785 cccgcacagcacgcgtccgcagcaccacaccccatctcgccctatgagcacctgc 4844
|||||
Db 133 CCCGCGACAGCACCGTCCCGGAGCACACCCACACCCCATCTCGCCCTATGAGCACCTGC 192

QY 4845 ttcgggcgtagtgcgtagacctgtatcgagccacatccccctggccttcgacccca 4904
|||||
Db 193 TTCGGGGCGTGAGTGGGTGAGACCTGTATGCGAGCCACATCCCCCTGGCCTTCGACCCCA 252
```



```

|||||
Db 89891 TGCTGAAACCAAGTCAGCGCGAGGTGGCGGACAAAAGGCCAGGTGCGCGCTGGGGGA 89832
QY 8148 acggatgtccgagactgagctgttttttttcaacatcgttgccgacggtggaagg 8207
Db 89831 ACGGATGCTCCGAGACATGGGACTGTTTTTTTACACATCGTTGCCGACGCGTGGGAAGG 89772
QY 8208 aaaggcagatgaaatgatgtgtgtttacagggtatatttttgataccttcaatgaat 8267
Db 89771 AAAGGCAGATGTAATGATGTGTGTGTACAGGGTATATTTTGTATACCTTCATGAAT 89712
QY 8268 taattcagatgttttacgcaaggagacttaccagttacttactgctgtgctgttttga 8327
Db 89711 TAATTCAGATGTTTACCAAGGAAGGACTTACCAGATTTACTGCTGCTGCTGCTTTGA 89652
QY 8328 tctgttaccgttccagagcggtgtgacgagcgacagtcggtgagcccaatcactcgca 8387
Db 89651 TCTCTGCTTACCGTTTCAAGAGCGGTGTGACGCGGACACATCGGTGACCCATCATCTGCA 89592
QY 8388 ggaccaaggggcgagactgctcg-tcacgccccgctgtctccctccctccctcc 8446
Db 89591 GGACCAAGGGGCGGGGACTGCTGCTCACGCGCGCTGTCTCTCCCTCCCTCCCTCC 89532
QY 8447 ttgggcagaatgaattcagtcgtattctgtggcgccatttgcgcaggtggtggtatt 8506
Db 89531 TTGGCAGCAATGAATTCGATGCGTATCTGTGGCGGCATCTGCGCAGGTTGGTATT 89472
QY 8507 ctgtcatttacacagtcgttcttaattaaaagggaattactcccaaaaaa 8561
Db 89471 CTGTCAATTACACAGCTGCTTCTAATTAAAGGCAATTTACTACAGTTACAA 89417

RESULT 10
AC027706 AC027706 161970 bp DNA HTG 02-MAY-2000
LOCUS Homo sapiens chromosome 12 clone RP11-665C13 map 12, WORKING DRAFT
DEFINITION
SEQUENCE, 49 unordered pieces.
AC027706
AC027706.2 GI:7677821
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 161970)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 12, clone RP11-665C13
Unpublished
2 (bases 1 to 161970)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,
Boguslavskiy, L., Boukhvalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, P., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Glinde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Melidrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severly, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
JOURNAL
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

COMMENT

On May 2, 2000 this sequence version replaced gi:7382510.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
 Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6713

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 140943 bases at least Q40

Consensus quality: 150450 bases at least Q30

Consensus quality: 154599 bases at least Q20

Insert size: 165000; agarose-fp

Insert size: 157170; sum-of-contigs

Quality coverage: 3.1 in Q20 bases; agarose-fp

Quality coverage: 3.3 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 49 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1052: contig of 1052 bp in length

1053 1152: gap of 100 bp

1153 2378: contig of 1226 bp in length

2379 2478: gap of 100 bp

2479 3812: contig of 1334 bp in length

3813 3912: gap of 100 bp

3913 5368: contig of 1456 bp in length

5369 5468: gap of 100 bp

5469 6519: contig of 1051 bp in length

6520 6619: gap of 100 bp

6620 7852: contig of 1233 bp in length

7853 7952: gap of 100 bp

7953 8464: contig of 512 bp in length

8465 8564: gap of 100 bp

8565 10347: contig of 1783 bp in length

10348 10447: gap of 100 bp

10448 12145: contig of 1698 bp in length

12146 12245: gap of 100 bp

12246 14006: contig of 1761 bp in length

14007 14106: gap of 100 bp

14107 15358: contig of 1252 bp in length

15359 15458: gap of 100 bp

15459 17152: contig of 1694 bp in length

17153 17252: gap of 100 bp

17253 18800: contig of 1548 bp in length

18801 18900: gap of 100 bp

18901 20291: contig of 1391 bp in length

20292 20391: gap of 100 bp

20392 21889: contig of 1498 bp in length

21890 21989: gap of 100 bp

21990 24355: contig of 2366 bp in length

24356 24455: gap of 100 bp

24456 26757: contig of 2302 bp in length

26758 26857: gap of 100 bp

26858 28317: contig of 1460 bp in length

28318 28417: gap of 100 bp

28418 30781: contig of 2364 bp in length

30782 30881: gap of 100 bp

30882 32789: contig of 1908 bp in length

32790 32889: gap of 100 bp

32890 35214: contig of 2325 bp in length

Qy	7574	gggaggcggtgdeuersperlugtcaggtccacgagcaccagaaacgcctgcagga	7633
Db	60639	GGCGGGGGGGGGGCGGTGCAGGTCCACGAGCACCAGAACGGCCCTGCAGGA	60698
Qy	7634	gcggggcggtgcgcgactcccccaaacgaagaaagagcccttgagtccgcctgcctcc	7693
Db	60699	GCAGGGCGGTTGCCGACTCCCCCAACCAAGAAGAGAGCCCTTGAGTCGCGCTCGCCCTCC	60758
Qy	7694	atccatctgtccgtccagagcgcgcatacttgcctgtctaagaaccttaactaagaactccc	7753
Db	60759	ATCCATCTGTCCGTCCAAGAGCGGCATCCTTGCCTGTCTAAAGCOTTAACTAAGACTCCC	60818
Qy	7754	qcccgggctggccctgtgcagacttactcaggggatgtttactcgtgctcgcggaag	7813
Db	60819	GCCCGGGCTGGCCCTGTGCAGACTTACTCAGGGGATGTTTACTCTGTGCTCGGAAGG	60878
Qy	7814	gaggggaagggcggggagggggcacgagcggtgtggcagccacacacagcgcgcca	7873-
Db	60879	GAGGGGAAGGGCGCGGGAGGGCACGGCAGGCGGTGTGGCAGCCACACGACGAGGGGCCA	60938
Qy	7874	.gggcggccagggaacccaaaagcagatgaccacgcacctccacgcactcgcctcccccgaa	7933
Db	60939	GGCGGGCCAGGGACCAAAGCAGGATACCAACGACCTCCACGCGACTGCCTCCCCCGAA	60998
Qy	7934	tgcatttgaaacaaagtctaaactgagctgcgaagccccgcgcctccctccgcctccc	7993
Db	60999	TGCATTTTGGAACCAAGTCTAACTGAGCTCGCAGCCCCCGCGCCCTCCCTCCGCGCTCCC	61058
Qy	7994	atcccgcttagcgcctctggacagatgaacgcagagcccttgtccagcccccatgcgcctgt	8053
Db	61059	ATCCCGCTTAGCGCTTGGACAGATGGACGAGGCCCTGTCCAGCCCCCATGTGGCGCTCGT	61118
Qy	8054	tccggtccccacagactgccccagccacacagattgctggaaaaacaagtccagggccaggtg	8113
Db	61119	TCCGGTCCCCACACACTGCCCCACGCCACAGAGATTGTGGAAAACAAGTCAGGCCAGGTG	61178
Qy	8114	ggcggaacaaaggccaggtgcgcctggggggaacggatgctccgaggactggactgtt	8173
Db	61179	GGCGGACAAAAGGCCAGGTGCGCCTGGGGGAACGGATGCTCCGAGGACTGGACTGTT	61238
Qy	8174	ttttcacacatcgtgccgcagcgggtgggaagaaaggcagatgtaaatgatgtgtcgg	8233
Db	61239	TTTTTTCACATCGTTGCCGAGCGGTGGGAAGGAAAGGCAGATGTAATGATGTGTTGG	61298
Qy	8234	tttacagggatatattttgatcacctcaatgaattaattcagatgttttacgcgaaggaag	8293
Db	61299	TTTTACAGGGTATTTTTTGATACCTCAATGAATTAATTACAGATGTTTATCCGCAAGGAAG	61358
Qy	8294	gactaccagtatctactcgtcgtcgttttgccttctcctcgtctaccgttcaagagcggtg	8353
Db	61359	GACTTACCCAGTATTACTGCTGCTGTGCTTTTGATCTCTGCTTTACCGTTCAGAGGCGGTG	61418
Qy	8354	tgcagggccagactcgggtgaccccatcactgcagggaccaaagggcgaggtactgcctcq-	8412

```

source
1. .161970
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/map="12"
/clone="RP11-665C13"
/clone_lib="RPC1-11 Human Male BAC"
1. .1052
/note="assembly_fragment"
1153. .2378
/note="assembly_fragment"
2479. .3812
/note="assembly_fragment"
3913. .5368
/note="assembly_fragment"
5469. .6519
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```


Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7063

Center clone name: 339_B_19

* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1
* 965 1064: gap of 100 bp
* 1065 2044: contig of 980 bp in length
* 2045 2144: gap of 100 bp
* 2145 3096: contig of 952 bp in length
* 3097 3196: gap of 100 bp
* 3197 4131: contig of 935 bp in length
* 4132 4231: gap of 100 bp
* 4232 5218: contig of 987 bp in length
* 5219 5318: gap of 100 bp
* 5319 6322: contig of 1004 bp in length
* 6323 6422: gap of 100 bp
* 6423 7388: contig of 966 bp in length
* 7389 7488: gap of 100 bp
* 7489 8466: contig of 978 bp in length
* 8467 8566: gap of 100 bp
* 8567 9527: contig of 961 bp in length
* 9528 9627: gap of 100 bp
* 9628 10612: contig of 985 bp in length
* 10613 10712: gap of 100 bp
* 10713 11683: contig of 973 bp in length
* 11686 11785: gap of 100 bp
* 11786 12725: contig of 940 bp in length
* 12726 12825: gap of 100 bp
* 12826 13794: contig of 969 bp in length
* 13795 13894: gap of 100 bp
* 13895 14875: contig of 981 bp in length
* 14876 14975: gap of 100 bp
* 14976 15963: contig of 988 bp in length
* 15964 16063: gap of 100 bp
* 16064 17033: contig of 970 bp in length
* 17034 17133: gap of 100 bp
* 17134 18122: contig of 989 bp in length
* 18123 18222: gap of 100 bp
* 18223 19199: contig of 977 bp in length
* 19200 19299: gap of 100 bp
* 19300 20275: contig of 976 bp in length
* 20276 20375: gap of 100 bp
* 20376 21340: contig of 965 bp in length
* 21341 21440: gap of 100 bp
* 21441 22409: contig of 969 bp in length
* 22410 22509: gap of 100 bp
* 22510 23478: contig of 969 bp in length
* 23479 23578: gap of 100 bp
* 23579 24528: contig of 950 bp in length

* 24529 24628: gap of 100 bp
* 24629 25614: contig of 986 bp in length
* 25615 25714: gap of 100 bp
* 25715 26685: contig of 971 bp in length
* 26686 26785: gap of 100 bp
* 26786 27766: contig of 981 bp in length
* 27767 27866: gap of 100 bp
* 27867 28836: contig of 970 bp in length
* 28837 28936: gap of 100 bp
* 28937 29940: contig of 1004 bp in length
* 29941 30040: gap of 100 bp
* 30041 31073: contig of 1033 bp in length
* 31074 31173: gap of 100 bp
* 31174 32163: contig of 990 bp in length
* 32164 32263: gap of 100 bp
* 32264 33255: contig of 992 bp in length
* 33256 33355: gap of 100 bp
* 33356 34343: contig of 988 bp in length
* 34344 34443: gap of 100 bp
* 34444 35421: contig of 978 bp in length
* 35422 35521: gap of 100 bp
* 35522 36507: contig of 986 bp in length
* 36508 36607: gap of 100 bp
* 36608 37558: contig of 951 bp in length
* 37559 37658: gap of 100 bp
* 37659 38645: contig of 987 bp in length
* 38646 38745: gap of 100 bp
* 38746 39754: contig of 1009 bp in length
* 39755 39854: gap of 100 bp
* 39855 40828: contig of 974 bp in length
* 40829 40928: gap of 100 bp
* 40929 41925: contig of 997 bp in length
* 41926 42025: gap of 100 bp
* 42026 43035: contig of 1010 bp in length
* 43036 43135: gap of 100 bp
* 43136 44097: contig of 962 bp in length
* 44098 44197: gap of 100 bp
* 44198 45179: contig of 982 bp in length
* 45180 45279: gap of 100 bp
* 45280 46245: contig of 966 bp in length
* 46246 46345: gap of 100 bp
* 46346 47341: contig of 996 bp in length
* 47342 47441: gap of 100 bp
* 47442 48397: contig of 956 bp in length
* 48398 48497: gap of 100 bp
* 48498 49441: contig of 944 bp in length
* 49442 49541: gap of 100 bp
* 49542 50510: contig of 969 bp in length
* 50511 50610: gap of 100 bp
* 50611 51575: contig of 965 bp in length
* 51576 51675: gap of 100 bp
* 51676 52628: contig of 953 bp in length
* 52629 52728: gap of 100 bp
* 52729 53677: contig of 949 bp in length
* 53678 53777: gap of 100 bp
* 53778 54756: contig of 979 bp in length
* 54757 54856: gap of 100 bp
* 54857 55849: contig of 993 bp in length
* 55850 55949: gap of 100 bp
* 55950 56913: contig of 964 bp in length
* 56914 57013: gap of 100 bp
* 57014 57990: contig of 977 bp in length
* 57991 58090: gap of 100 bp
* 58091 59058: contig of 968 bp in length
* 59059 59158: gap of 100 bp
* 59159 60147: contig of 989 bp in length
* 60148 60247: gap of 100 bp
* 60248 61223: contig of 976 bp in length
* 61224 61323: gap of 100 bp
* 61324 62301: contig of 978 bp in length
* 62302 62401: gap of 100 bp
* 62402 63354: contig of 953 bp in length
* 63355 63454: gap of 100 bp

[illegible]

[illegible]

Search completed: September 07 2007 15:13:12
Job time: 13279 sec

```

RESULT 15
HSU80761/c
LOCUS
DEFINITION Homo sapiens CTG26 alternate open reading frame mRNA, complete cds
ACCESSION U80761
VERSION U80761.1 GI:2565090
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 560)
AUTHORS Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
TITLE CDNAs with long CAG trinucleotide repeats from human brain
JOURNAL Hum. Genet. 100 (1), 114-122 (1997)
MEDLINE 97369492
REFERENCE 2 (bases 1 to 560)
AUTHORS Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of
Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA
FEATURES
source
1..560
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="fetus"
/tissue_type="brain"
<1..560
/gene="CTG26"
<1..375
/gene="CTG26"
/notes="cysteine rich"
/codon_start=1
/product="CTG26 alternate open reading frame"
/protein_id="AAB91452.1"
/db_xref="GI:2565091"
gene
CDS

```


31-MAR-1999; 99US-0127607.

-Qy 1500 ERACYEESLSKRPCTASSSGGSIARGAPVIVPELKGKPRQSPBLYEDHGAPFAGHLPRGSP 1559
Db 1501 eracyeeslkrpctasssggsiargapvivpelkgqrspblyedhgapfaghlprgsp 1560
Qy 1560 VTMREPTERLOEGSLSSSKASQDRKLTSTPREIAKSPHSTVPEHHPHPISEYEHLLRGVS 1619
Db 1561 vttrptprlqegslssskasqdrkltstpreiakspstvphehphpispseyehllrgvs 1620
Qy 1620 GVDLYRSHIPAFDPTSIPIRGIPDLAAAYYLPRHLAPNPTYPHLYPPYLIRGYPDTAAL 1679
Db 1621 gvdlyrshipafdptsipirgipdlaaaayylprhlapnptyphlyppylirgypdtaal 1680
Qy 1680 ENRQTIINDYITTSQMHNTATAMAQRADMLRGLSPRESSIALNYAAGRGIDLSQVPH 1739
Db 1681 enrqtiindyitstqgmhntatamaqramlrglspressialnyaagrgidlsqvp 1740
Qy 1740 LPVLVPTPTGPTAMADRLAYLPTAPOPFSSRHSSPLSPGPHLTLPKPTTSSSERERD 1799
Db 1741 lpvlvptptgpcacamlrlaylptapqfssrhssplspgphltltpkpttsssererd 1800
Qy 1800 RDRERDREREKSLTSTTTVEHAPITWRPGTEQSSGSSGGGSSSRPASHSHAHQ 1859
Db 1801 rdrerdrekeksltstttvehapitwrpgteqssgssggsssrpashshahq 1860
Qy 1860 HSPISPTQDALQORPSVLHNTGKGIITAVEPSKPTVLRSTSSVPRPAATPPATHC 1919
Db 1861 hspisprtqdalqprsvlhtgmkgiitavepstvlrststssvprpaatppathc 1920
Qy 1920 PLGCTLDGVPTLMEPVLLPKAEAPRVARPERPRADTGHAFKAPKPSGLESPPSPKGS 1979
Db 1921 plgctldgvptlmevllpkaeaprvarperpradtgghafkpparsgglepasspakgs 1980
Qy 1980 EPRPLVPVSGHATIAITPAKNLAPHASDPDPAPPASADPHREKTSQKPFESIQEELR 2039
Db 1981 eprplvpvsghatiartpaknlaphasdpdpappasasdphektqskpfesiqelr 2040
Qy 2040 SLGYHSGSYSPGVEPVSPVSSSLTHDKGLPKHLELDKSHLEGELRKPQGPVKLGE 2099
Db 2041 slgyhsgsysegvepvspvssslthdkglpkhleeldkshlegelrkpgpvyklge 2100
Qy 2100 AAHLPHLRPLPESOPSSPLLOTAPGVKGHQRVVTLAQHISEVITQDTRHHPQOOLSNPL 2159
Db 2101 aahlphlrplpesopsspllotapgvkgqrvvtlaqhisevitqdytrhhpgqlsapl 2160
Qy 2160 PAPLYSPFGASCPVLDLRRPDSLYLPPPHGAPARGSPHSEGGKRGPEPKNTSVLGGGE 2219
Db 2161 paplyspfgascvldlrrpsdlylppphgaparpsphseggkrpepkntsvl99ge 2220
Qy 2220 DGIPEVSPPEGMTEPHGHSRAVYPLLYRDGEQTEPSRMGSKSPGNTSQPPAFFSKLTESN 2279
Db 2221 dgievsppegmtpeghsraavypllyrdgeqtepsrmgskspgntsqppaffskltesn 2280
Qy 2280 SAMVSKKQELINKLNTHNRPEYNTSQPCTEFNPAITGTGLMYRQAOVEHASTN 2339
Db 2281 samvskkqelinklnchnrpeynlsqpgteifnmpaitgtglmyrqaqvqehastn 2340
Qy 2340 MGLBAILKALMGKYDQWESPPLSANAFNPNASASILPAAMPITAADGRSDHFLTSPGG 2399
Db 2341 mgleailkalmgkydqwespplsanafnnpnasasilpaampitaadgrsdhfltspgg 2400
Qy 2400 GGAkakvSGRPSRRKAKSPACGLAGDRPPSVSVHSGDCNRRRPLNVRVWEDRPSAGS 2459
Db 2401 ggakvsgprpsrrkakspaglagdrppsvsvshsgdcnrrrpltnrvwedrpssags 2460
Qy 2460 TPPYPNPLMRLOAGVNASPPPCPLPAGSPLAGPHAWDEEPKPLCSQYETLSDSE 2517
Db 2461 tppypnplmrloagvmaspppcplpagsplagphawdeepkplcsqyetlisdse 2518

RESULT 2
AAW18226
ID AAW18226 standard; Protein; 1495 AA.
XX

AC AAW18226;
XX 24-SEP-1997 (first entry)
DT
XX Transcriptional co-repressor SMRT.
DE
XX
XX Silencing mediator for retinoic acid and thyroid hormone receptor;
KW SMRT; transcriptional co-repressor.
KW
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
XX Region 1..160
FT /label= N-terminal_region
FT /note= "proline-rich domain"
FT
FT Region 773..790
FT /label= ERDR_region
FT Region 812..827
FT /label= SG_region
FT Region 1061..1132
FT /label= glutamine-rich region
FT Region 1201..1495
FT /label= C-terminal_region
FT Peptide 1330..1376
FT /note= "alternatively spliced insert not
FT present in the original two-hybrid
FT clone"
XX
PN M09709418-A1.
XX
XX 13-MAR-1997.
PD
XX
XX 24-JUL-1996; 96WO-US12371.
PF
XX
XX 01-SEP-1995; 95US-0522726.
PR
XX
XX (SALK) SALK INST BIOLOGICAL STUDIES.
XX
XX Chen JD, Evans RM;
PI
XX WPI; 1997-192894/17.
DR
XX
XX New co-suppressor of steroid-thyroid hormone receptor activity -
PT also methods for identifying compounds that relieve its suppressant
PT effect and/or activate receptors
PT
XX
PS Claim 2; Page 40-45; 7lpp; English.

XX
CC A novel receptor interacting factor (AAW18226) is designated SMRT,
CC i.e. silencing mediator (co-repressor) for retinoic acid receptor
CC (RAR) and thyroid hormone receptor (TR). Its association with RAR
CC and TR both in solution and on DNA response elements is
CC destabilised by ligand. The interaction of SMRT with mutant
CC receptors correlates with the transcriptional silencing activities
CC of receptors. In vivo, SMRT functions as a potent co-repressor. A
CC GAL4 binding domain fusion of SMRT behaves as a frank repressor
CC of a GAL4-dependent reporter. These data identify a novel class of
CC cofactor which is believed to represent an important mediator of
CC hormone action. Full-length cDNA for SMRT has been isolated from
CC a HeLa library in a two-hybrid screen using a GAL4 DNA binding
CC domain/RXR fusion protein as bait.
XX
SQ Sequence 1495 AA;

Query Match 59.5%; Score 7866; DB 18; Length 1495;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1486; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1031 DKEAFAAEAKLPGDPCWTSGLPFPVPPPREVIKASHADPPSAFSAFGCHPLPLGLHD 1090
Db 9 dkeafaaeklpdpcwtsglpfpvpppprevikasphadpsafsyappghplplghd 68

Qy	1091	TARVPLPRPTTINBPPLISSAKHPSVLBRQTGATISQMGSVOLHVYPSEHAKAPVGPVTM	1150
Db	69	tarpvlprptinbpplissakhpsvlerqigaismqsvqlhvpseyehakapvgpvtm	128
Qy	1151	GLPLPMDPKLAPFGVKQEOILSPRGOAGPPESLGVPTPAQEAASVLRGTLGSLVPGGSITK	1210
Db	129	gplpmpdklapisvkgdeqlsprgggppeslgvptaqaasvrlrgtaigsvpvgssitk	188
Qy	1211	GIPSTRVPSDAITYRGSTHGTTPADVLYKGTITRIIGEDSPSLRDGRSDSLPKGHVIV	1270
Db	189	gipstrvpsdaityrgsthtgtpadvlykgtitriigedspslrdgrsdslrpkghviy	248
Qy	1271	EGKKGHVLSYEGGMSVTVQCSKEDGRSSSGPPHETAAPKRTYDMMEGRVGRSAISSASIEGL	1330
Db	249	egkkghvlsyeggmsvtqcskedgrsssgpphetaapkrtydmmegrvgraisasiegl	308
Qy	1331	MGRAIPPERHSPHHLKEOHIIHIGSTIQGTIPRSYVQAQEDYLRRRAKLLKRGTPPPPPPS	1390
Db	309	mgraipperhsphhlkeqhnirgslcqqiprsyveaqedyllrreakllkregtpppppps	368
Qy	1391	RDLEAYKTOALGPLKLKPAHEGLVATVKEAGRSIHEIPREBELRHTPELPLAPRLKEGS	1450
Db	369	rdleayktcalgplklkpaheglvatvkeagrsiheipreelrhtpelplaprlkegs	428
Qy	1451	ITQGTPLKYDTGASTTGSKKHVDVRSILIGSPGTFPPVHPLDVMADARALERACVYESLSKS	1510
Db	429	itdgtplykdtgasttgskkhvdvrsiligsptgfppvhpldvmadalaracvyeeslks	488
Qy	1511	RPGTASSGGSGIARGAPVTVPELGPQRQSPLVYEDHGAFFAGHLPRGSPVTMRPTEPRLQ	1570
Db	489	rpgtcassgsgsiargapvtpvelgprqspvlvyedhgapfaghlprgspvumreptprlq	548
Qy	1571	EGSLSSSKASQDRKLKLTSTPREIAKSPHSVNPBHPHPISPYEHLLIRGVSQVDLYRSHIPL	1630
Db	549	egslssskasqdrklkltstpreiakspbstvpehhphispvyeihllirgvsqvdlyrshipl	608
Qy	1631	AFDPTSTPGIPLDAAAAYLPLPHLAPNPTYPHLYPPYLIRGYDPTAALENRQTIIINDYI	1690
Db	609	afdpstprgipldaaaayvlphlapnptyphlyppylirgydpdtaalenrqtiindyi	668
Qy	1691	TSQOMHHNTATMAQADMRLGLSPRESSIALNYAAGPRGLIDLSQVPHPLVLPVPPPGT	1750
Db	669	tsqomhntatamaqadmrlglspressialnyaagprglidlsqvpplvlpvpppgt	728
Qy	1751	PATAMDRLAYLPTAPQPFSSRSSSSPLSPGGPTHLTKTPTTSSSERERDRDRDRDRER	1810
Db	729	patamdrlaylptapqpfssrhsspslpggpthlktpttssererdrdrdrdrer	788
Qy	1811	EKSLTSTTTTVEHAPIWRPGETBOSSGSGSGGGGGSSRSPASHSHAHQHSPISTPRTQDA	1870
Db	789	eksiltstttvehapirwpgteqsgsgsgsgggsgssrpsashshahqhspsiprtqda	848
Qy	1871	LQORPSVLHNTGMKGIIITAVEPSKPTVLRSTSTSPVRPAATFPFPAHCHPLGGTLDGVYP	1930
Db	849	lqdrpsvlhntgmkgiitavepskptvlrststspvrpaatfpfpahtcplggldgvyp	908
Qy	1931	TLMEPVLLPKAEPRVARPERPRADTGHAFLAKPPARSGLEPASSPSKGSSEPRPLVPPVSG	1990
Db	909	tlmepvllpkaeprvarperpradtgghaflakpparsglepasspskgsseprplvpvsg	968
Qy	1991	HATARTPAKNLAPHASDPDPAPPASADPHREKTQSKPFSIQEELRSLRGYHGSSYSP	2050
Db	969	hatiartpknlaphasdpdpappasadphrektqskpfsiqeelsrlsyhgssysp	1028
Qy	2051	EGVEPVSPVSSPSLTHDKGLPHLELDKSHLEGLRPKQCPVKLGGEEAHLPHLRPLP	2110
Db	1029	egvepvspvsspslthdkglphleldkshleglrpkqcpvklggeeahlphlrplp	1088
Qy	2111	ESOPSSSPILLQTAGPVKGHORVVTLAQHTISEVITQDYYTRHHHPQOLAPLPAFLYSPFGAS	2170
Db	1089	esqpssspillqtagpvkgqhorvvtlqhlisevitqdytrhhpqolsaplpalyfspgas	1148
Qy	2171	CPVLDLRPPPSDLYLPPPDHGAAPARGSPHSEGGKRSPEPNTSVILGGEDEGIEPVPSPPEG	2230

Db	1149	cpvldlrpsdlylppdhgaparqshsegkrspepnktsvlggedgiepvsppeg	1208
Qy	2231	MTPEGHSRSAVYPLLYRDGEOTEPSRMGSKSPGNTSQPAFFSKLTESNSAMVSKKQEI	2290
Db	1209	mtepghsarsavypllyrdgeqtepsrmgskspgncsqppaifskltesnsamvskkqei	1268
Qy	2291	NKkLTHNRNEPEYNISQPGTEIFENMPAITGTGLMTYRSQAVQEHASTNMGLEAIIRKAL	2350
Db	1269	nkkLthnrnepeynisqpgteifnmpaitgtglmtysqavqehastnmgleaiirkal	1328
Qy	2351	MGKYDQWEESPPLSANAFNPLNASLPAAMPITAADGRSDHTLTSPCGGGKAKVSGRPS	2410
Db	1329	mgkydqweespplsanafnplnasalpaampitaadgrsdhtltspggggkakvsgrps	1388
Qy	2411	SRKAKSPAPGLAGDRPPSVSVHSEGCNRRTPLTNRWEDRPSSAGSTPPYNPLIMR	2470
Db	1389	srkaksapglagdrppsvsvhsegdcnrrtpltnrwedrpssagstppynplimr	1448
Qy	2471	LQAGVMASPPPGPLPAGSGPLAGPHHAWDEEPKPLLCQYETLSDSE	2517
Db	1449	lqaymaspppgplpagsgplagphhawdeepkpllcqyetlsdse	1495
RESULT 3			
AAB12454			
ID	AAB12454 standard; Protein; 2453 AA.		
AC	AAB12454;		
DT	24-OCT-2000 (first entry)		
DE	HNRCR protein sequence.		
KW	Human; HNRCR; nuclear receptor coreceptor.		
OS	Unidentified.		
PN	CN1250094-A.		
PD	12-APR-2000.		
PF	06-OCT-1998; 98CN-0120919.		
PR	06-OCT-1998; 98CN-0120919.		
PA	(XINH-) XINHUANGPU FUDAN GENE ENG CO LTD SHANGHA.		
PI	Yu L, Tu Q, Zhao Y;		
DR	WPI; 2000-400830/35.		
DR	N-PSDB; AAA60630.		
PT	Preparation of new human kernon acceptor co-repressor coding series and the polypeptide -		
PS	Example 2; Fig 2; 58pp; Chinese.		
CC	The present invention describes a human homologue of nuclear receptor coreceptor (HNRCR). The present sequence represents an HNRCR protein sequence used in comparison with the human HNRCR.		
SQ	Sequence 2453 AA;		

Query Match	31.7%	Score 4187;	DB 21;	Length 2453;
Best Local Similarity	41.0%;	Pred. No. 1.6e-277;		
Matches 1103;	Conservative 348;	Mismatches 796;	Indels 444;	Gaps 106;
Qy	16	EPBRYPPHSLSPVQIARTHTDVGLELY--HHSDRYASHLSPGSIIPQRRRPSELLSEFQ	73	
Dd	17	qsrypsnaqvtyfpssatqhqqefaydvrrshllevsqslqlqqqqqqqlrrlpallsefh	76	

QY 74 PGNERSQELHLRSHESHYLPGLGKSMETIESKRPRLELLPDLR-----PSPLIATQ 128
Db 77 pgsdrpqe---rrsgyeqfpgpsvdpdhdshleskrprleqvsdshfgrisaavplvht-1 133
QY 129 PAG---SEDLTKDRSLTGKLE-PVSPSPPHWTDPELELYPPRLSKELTONMDRVOREIT 184
Db 134 peglrssanakdpafgvkhespsplsgqpcgddqnaspsklskeeligsmrdvrdela 193
QY 185 MYEQOISKLKKOOLLEEAAPPEPEKVPSPPTIESKRSRLOIYIDENRKKAEAAHRI 244
Db 194 kveeqilkkkqggleeaaakppekvpvppveqkhrsvqilvdenrkkaeaaehk1 253
QY 245 LGLGQVVELYINOPSDTROQHENIKINQAMRKKLILYFKRNHARKQWKQFCORYDQ 304
Db 254 feglgpkvelynpsdtkvvyheniktqvmrkkllffkrnrharkqrechkicqrydq 313
QY 305 LMEALEKKVERLENNPRRAKESKREYIEYKOFPEIRKORELOERMOSRVGQSGLSMS 364
Db 314 lmeaewekvdriennprrrakeskrtreyekqfpeirkqredqerfq-rvqgrgaglsat 372
QY 365 AARSEHVESEIIDGLSEQENLEKQROLAVIPPMYLDADOQRIKFINNGLMADPMKYK 424
Db 373 iarsheliseidglseqqennekqmrqisvipmmfdacqrvkvkfinnmglnmedpmkyk 432
QY 425 DROVNMWSEQKETFREKFMQHPKNFGLIASFLERKTVAECVLYYLTCKNENYKSLVR 484
Db 433 drqfmvvtdeheikfkdkfidhpknfgliaasylerksvpcvlyytlcknennykalvr 492
QY 485 RSY-RRRKSQOQOQOQOQOQOQOQOQPMPSRQSEKDEKEKEAEK--EEEKPEVEND 541
Db 493 rnygrrngqgarpsequeekveek-----eedkaektekkeekdkdeekddeds 544
QY 542 KEDLLKETDDTSGEDNDEKEAVASKGRKTANSQGRKRITRSMANE---ANSEEAITP 598
Db 545 kettkekrdrateapeereeqytrgrktansqgrkvtrtsmseaaaaanaaaate 604
QY 599 QOSAEI-----ASMELENESSRWTEEMETAKKGLLEHGRNWSAIARMVGSKTVSQCKNF 652
Db 605 eppplpppepistepvetsrteeemevakkgivehgrnwaaiakmvgtkseaqcknf 664
QY 653 YFNYYKRONLDILQOHLKMEKERNARRKKKAPAAASEEAFFPVVEDEMEASVSG 712
Db 665 yfnyrrhmlnldqhkqarkpreerdsvqcesvastsa---qededieas--- 716
QY 713 NEEEMVEEAALHSGNEVPRGECGPATVNNSSDTESTIPSPH-TEAAK---DTQONGPK 768
Db 717 neeneppedseg-----aensdtesapspsveaakssedsena-- 756
QY 769 PPATLGADGPPGPTPPRTSRAPTEPTPASEATGATPPPPAPSPSPAPPVVPKBEKE 828
Db 757 -----asrgntepvaeleatdpapcasp--ssavpttkpaeres 794
QY 829 ER-----TAAAPPVEGEO-----KPPAAEELAVDTCKAEPEVKSECTEAE 871
Db 795 veaqvtdasaetaepmdvheecqaegssvldppapck--adsvdpemqv-pentaskg 851
QY 872 EGPAKGDAEAAEATAEGALKAEKKEGSGRATTAKSSGAPO---DSDSATFCSADEYDE 928
Db 852 egdakerdlest-----sekteardedvvvvaerpepqsddssatcsadegvd 902
QY 929 AEGGDKNL--LSPRPSLLTPTGDPANASQPK--LDLKQLKQRAAAIIPPI----- 976
Db 903 ge-perqrvfmdakpalsltppgsali--sslpknalldlpqlqhraavipmvsvctpcni 960
QY 977 -----QVTKVHPPPREDAAPTAPAPPPQNLOPESDAPQPGSSPRGK 1022
Db 961 pigtpvsvyalyqrhikamhesalle-----eqrqrgeqvdecrsscspest 1008
QY 1023 SRSPAPPADKAEFAAEAKLPGDPPCWTSGLPFPVPPPIKASHPADPSAFSVAPPCH 1082
Db 1009 sksp-----nre-----w-----evlqpaph-----qvntnlpegv 1034
QY 1083 PLPLGLHDTARVLPRLPPTISNPPLISSAKHPSVLERIGAISQG-----MSVOLHPY 1137

Db 1035 rlp-----cttrtrppplipsskttvasek-psfi--mggsisqqtptgtyishnqavp 1086
QY 1138 SEHAKAPGVMTGLPLPMDPKKLAPFSGVKQQLSPRGOAGPPEPSLGVPVTAQEAASVLRG 1197
Db 1087 qeapkpvsqgisliglprqquestkaapltlykgeefsprsqnsqegllv-rachegvvrg 1145
QY 1198 TALGSPVGSITKIPSTRVPDSAITYRGSITHTGPPA-----DVLVYKGTIRI-IG 1248
Db 1146 ta-gavqegsilrtgtpaskisvetisslrsgsitqtpalpqaqigpteaalvgkpvsmple 1204
QY 1249 EDSPSRLDRGRDSDLPKGHVIEGKGHVLSEGGMSVTCQSKEDGRSSGPPHETAAPK 1308
Db 1205 esspekv---reeaaskghviegkshilsydnikna-----regtrsprtahems1-k 1255
QY 1309 RYDMMEGRVGRASIS-----SASIEGLMGRAPPRHRSPH-HLKEQHHRIGSITGIPR 1361
Db 1256 rseyavegsikgmsmrespsvaapleglicralp--rgspshdtkertvlsqsmgtpr 1313
QY 1362 SYVEAQEDYLRREAKLLKREGTPPPPPSRDLTEAYKTOALGPLKLKPAHEGLVATVKEA 1421
Db 1314 ataesfedgl-kypqikres-----ppirafegai-----tkgkp-ydg-ittikem 1358
QY 1422 GRSIHEIPREEL-----RHTPELPLAPRLKEGSGITQGTPLKYDYTGASTGSKKHDRVSL 1476
Db 1359 grsiheiprqdiltqesrktpevvqstrpiegsisqgtplkfdnn-sqgsaikhnvksl 1417
QY 1477 ICSPTGTPPVHPLDMAD-ARALERACYE-----ESLSKRPGTASSGSGSTARGAPVIV 1530
Db 1418 itgpskl--prgmleivpenikvergkyedvkaqevrartsvvsgpsvlrst---1 1472
QY 1531 PELGHPROSPLRYEDHGA-----PEAGHLPRGSPVTMREPTPRLQBGSLSSSK-ASODRK 1584
Db 1473 heapkaqlspglydassarrtpvsgntisrsgpmnmr-----tsdvssksasherk 1525
QY 1585 LNSTPRE-----TAKSPHSTVPEHHPHPIPSYEHLLRGVSDLYKSHIPLAIFDPTSPRG 1640
Db 1526 sltltqgresipakapvgvqdvivsh--spfqphrseaaagevy-shlpthldp-ampfh 1582
QY 1641 IPLDAAAAYLLPRHLAPNTYPHLYPPYLIRGYPDTALEN-RQIINDYITSOQHHT 1699
Db 1583 raidpaaayllqrlsptpgypsgvlyq-----amentrltilndyitsqgmval 1633
QY 1700 ATAMAQADMLRGLSPRESSLALNYAAGPRGIIDLQVPHLPVLPVPTPGTATAMDRLA 1759
Db 1634 -----rpdvtrglspreqplgypa-trgiidltmnp-tilvphaggtstppmdrit 1685
QY 1760 YLPTAPQPFSSR-HSSPSLSPGGPHTLTKPTTSSERDRDRDRDRDREREKESILTST 1818
Db 1686 yipgtqvtfpprpynaaslsqgphthl---aaaaaererererekerereter 1742
QY 1819 TTVEHAP---IWRPGTEQSSGSGSGSGSGSGSRPASHSHAHQHSPTSPRQDA-LQOR 1874
Db 1743 eriaaapadlylrpseque-----pgrpgshgyvrspsp-svrtqetlqgr 1787
QY 1875 PSVLHNTGMGIITAVEPSKPTVLRSTSTSSPV-----RPAATFPFATHPLCGTLDGVVP 1930
Db 1788 psvfgtngtswitpldptaqrlimplpsggpsisqglpaasyntaada-laalvdaas 1846
QY 1931 TLMEVPLLPKEAPR-----VARPERPRADTGHAFKAPPAKSGLEPASPSPSK 1977
Db 1847 apqmdvskteskshaarleenlrarsaavseqqlqeknllevkrsvqvcvtssalpsg 1906
QY 1978 GSEPRPLV-----PPVSGHATTIARTPAKN-LAPHHASPPDPAPPASADPHRE 2024
Db 1907 kaqphasvvyseagkdkpppkryeelertrgkttitaanfildvlttrqiasdkdar 1966
QY 2025 KTQSPFSIQIELESLGSHGSSYSPGVEPVSPPSVSSSLTHDKGLPKHLELDKSHLEG 2084
Db 1967 gqsqsdss-----sslsshryetasdaievispasspappqekpqayqpdmvaqnaen 2020
QY 2085 ELRPQKQPGVPLGGBAHLPHLR-----PLPESQSPSSPLLQT---APGVKQHQRVVTLAQ 2137

Db 2021 estrgyegp-----lhhyrsqgespsqgqplppssqsegmgqvrthrliatd 2071
Qy 2138 HISEVITQYTRHH-PQOLSAPL-----DAPLYSFFGASCAPVLDRRPPSDLYLPPPD--- 2189
Db 2072 hicqitqdfanqpsqaststftspalstp-----vrtktsryspesqst 2123
Qy 2190 --HGAPA-RGSPG-----SEGKRSPEPNKTSVLGGEDGIEPVPSPGEMTEPG-HRSRA 2240
Db 2124 vlhprpgprvspenlvdkargprgkspershi---pspeypespqg---pavhekqd 2177
Qy 2241 VYPLLYRDEQTEPERMSKSPCNTSQPPAFPSKLTESAMVSKKOEINKKLTHNRN 2300
Db 2178 smlllsgrvpaedrsdsrpsyslpfctkl-estspmvkakkqefrklsssgg 2236
Qy 2301 EPEYNISOGTEIFNMNATGTGLMTYRSQAOEHASTNMGLEAIRKALMGKYDQWEE- 2359
Db 2237 dsdmaaagqgteifnlpavttsgavssrshsfadpas-nigledlrkalmgfsfdkved 2295
Qy 2360 -----SPPLSANAFNLNASASLPAAMPITAADGRSDHTLTSPGGG-GKAKYSGRPSRRK 2413
Db 2296 hgvnmshpv---gimpgaastav-----vtssearrdegepsphagvckpklinsknsrk 2347
Qy 2414 AKSPAPGLA--SGDRPPSVSVHSGDCNRRPPLNVRWEDRPPSAGSTPPPNPLIMRL 2471
Db 2348 skspipggsgylgterpsvsvhsegdyhrqtp--gwawedrpsstgstqfpynpplfirm 2405
Qy 2472 QAGVNASPPPPCLPAGSGPL--AGPH--HAWDEPKPLLCQYETLSDSE 2517
Db 2406 ----lsstpttiacapsaitaaphqgnriwerepapllsyaqecilsdsd 2452

RESULT 4

AA12453
ID AAB12453 standard; Protein; 2343 AA.
AC AAB12453;
XX
XX 24-OCT-2000 (first entry)
XX Human HNRCR protein SEQ ID NO:20.
XX Human; HNRCR; nuclear receptor coreceptor.
XX Homo sapiens.
XX CN1250094-A.
XX 12-APR-2000.
XX 06-OCT-1998; 98CN-0120919.
XX 06-OCT-1998; 98CN-0120919.
XX (XINH-) XINHUANGPU FUDAN GENE ENG CO LTD SHANGHA.
XX Yu L, Tu Q, Zhao Y;
XX WPI; 2000-400830/35.
XX N-PSDB; AAA60629.
XX Preparation of new human kernon acceptor co-repressor coding series and the polypeptide -
XX Claim 4; Page 25-30; 58pp; Chinese.
XX The present sequence represents a human homologue of nuclear receptor coreceptor (HNRCR).
XX Sequence 2343 AA;

Query Match 30.6%; Score 4043; DB 21; Length 2343;
Best Local Similarity 41.3%; Pred. No. 2.1e-219;

Matches 1061; Conservative 330; Mismatches 731; Indels 444; Gaps 104;
Qy 137 KDRSLTGLKLE-PVSPSPPHDTPLELVPPRLSKBELIQNMDRVDRITMTVQEQISKLLK 195
Db 36 qdpafggkheapspsisgqpcgddqnaspsklskeelqsgmdrvdrelakveqqlklkk 95
Qy 196 KQQLLEEAAPPEPEKPVSPPIESKHSLSVQIIYIDNRKKAEEAARILEGLGQVQELP 255
Db 96 kqqqlleeeaaapepekpvsppveqkhrsviqvliidenrkkaeeahkifeglgpkvelp 155
Qy 256 LYNQSDPTROTHENIKINQAMRKKLILYFKRRNHARKQWKOKFCORYDQLEALEKKYVER 315
Db 156 lynyqsdtkvynheniktnqvmrkkllifkrnrharkredqkicrydqlmeawekkvdr 215
Qy 316 IENNPARRAKESKVREYKOFPEIRKQRELQRMOSRGVORGSLMSAARSEHEVSEI 375
Db 216 iennprrrakesktrreyekqfpeirkqregqrfq-rvqrgaglsatarseheisei 274
Qy 376 IDGLSEQENLEKQMRQLAVIPPMGLYDADQORIKFTINMGLMADPMKYKDRQVNMWSEQ 435
Db 275 idglsegesnekqmrqlsvipmmmdaeqrkvkfimmgmedpmkvkykdrqfmvwdh 334
Qy 436 EKETPREKFMQHPKNGFLIASFLERKTVACVLYLYLTKKNENYKSLVRSY-RRRGSQ 494
Db 335 ekeifkdkfidhpnkngfliasylerksvpcvlylyltkknenykalyvrnygkrgrnq 394
Qy 495 QQQQQQQQQQQQQOOPMPSRQOEKDEKEKEKEKEEPEVENDEKLLKEKTDSTS 554
Db 395 qlarpsqeekeveeke--dkaektkekkekdeekdeekeskentke---kdkidgta 449
Qy 555 GEDNDEKEAVASKGRKTANSQGRKRGRITRSMANEEAEAITPOOS-----AEL 604
Db 450 -eeteereqatprgrktansqgrkrgritrsmtneaaaaaataateepppppp 508
Qy 605 ASMELNESSRWTEEMETAKKGLLEHGRNWSAIAIRWGSKTVSQCKNIFYNKKQNLD 664
Db 509 lstepvetsrteeeemevakkglvehgrnwaaiakmvgtkseagcknfyfnykrhndn 568
Qy 665 LIQOHKLMEKERNARRKKKAPAAASEEAFFPPVVEDEMEASGVSGNEEMVEAEAL 724
Db 569 llqhqkqtsrkpreedvsgcesvastva---qededieas---neenpedseg- 619
Qy 725 HASGNEVPRGECGPATVNNSSDTEIPSPH-TEAAKTGQNGPKPPATLGAAGDPPPPGP 783
Db 620 -----aenssdtesapspsveavk-----psed 643
Qy 784 TPRRTSRAPTEPTPASEATGATPPAP-PSPS-APPVVPVPEKEKEETEAAAPV---- 837
Db 644 spenatsrgntepavele-----pttetapstpslavpstkpaedesvetqvndisaet 699
Qy 838 -----EGEQKPPAAEELAVDTGKAEEPVKSECTEAEAGPKAGKDAEAAE 884
Db 700 aeqmdvdqgehsaeegsvcdppppack--adsvdvevrvpnenhaskvegdntkerldra- 756
Qy 885 ATAEGALKAEKKEGGSGRATTAKSSGA----PQ-DSDSATCSADEVDEAEQDKNRLLS 939
Db 757 -----sekreprdedlvvaqinaqrpeqsgndssatscade--dvdgepergrmf 806
Qy 940 P---RPSLLTPTGDPANASPOK--PLDLKQLKQRAAAIPPI----- 976
Db 807 pmdskpsllnptgslv--sslikpnpldlpqllqhraavippmvscstpcnpiptvpsgva 865
Qy 977 ---QVTKVHEPPREDAAPTKPAPPAPPQNLOPESDAPQOOPSGSRKSRSPAPPADK 1032
Db 866 lyqrhikamhesalle-----eqqrqeqidlecrrstscpgtsksp-----nr 909
Qy 1033 EAFRAEAQKLPDPPCWTSGLPFPVPPREVIKASPHADPPSAFSYAPPGHPLPLGLHDTA 1092
Db 910 e-----w-----evlqaph-----gvitnlpegvrlp-----tt 934
Qy 1093 RPVLPRPPTISNPPPLISSAKHPSVLERQIGAISOG-----MSVLQHVYPYSEHAKAP-VG 1146
Db 935 rtrpppppplipskttvasek-psfi--mgsisicgtptgtlytshnqasytqetpkpsvg 991


```

SQ Sequence 619 AA;
Query Match 5.4%; Score 718; DB 17; Length 619;
Best Local Similarity 36.0%; Pred. No. 1.5e-32;
Matches 213; Conservative 76; Mismatches 212; Indels 90; Gaps 26;
QY 1975 PSKGEPRPLV-----PPVSGHATIARTPAKN--LAPHASDPDPAPPASADSP 2021
Db 70 psgkqphasvvyseagdkgppksryeeelrgrtkttitaanfdivttrqlasdkda 129
QY 2022 HREKTSKPFISQIELESLRSLGVHGSYSPEGVPEVSPVSSPDLTHDKGLPKHLELDKSH 2081
Db 130 rergsqsdss-----sslsrhryetaadaievspassappqekpqayqpdmvkanq 183
QY 2082 LGEELRPQGPVKLGGEAAHPLHUR-----PLPESQSSSPLLQT--APGVKGHRVVT 2134
Db 184 aenestrgyegp-----lhvrsqdespspgqqlpsssqsegmgvprthrilit 234
QY 2135 LAQHTISEVITQDTRHHPOQLSAPLAPLYSPGA--SCPVLDLRRPPSDLYLPPD--- 2189
Db 235 ladhicqitqafarn--qvpqgststfqtspalsstpv---rtktsryspesqst 289
QY 2190 --HGAPA--RGSPPH-----SEGGKRSPEPNKTSVLGGGEGDIEPVSPPEGWTEPG--HSRSA 2240
Db 290 vlhprpgrvrspenlvdksgsrpksrsh---psepyepispgq---pavhekqd 343
QY 2241 VYPLLYRGEQTEPRSMGSKSPGNTSQPPAFPSKLTESNAMSVMKSKOEINKKLNTHNRN 2300
Db 344 smllsqrgvdpaedqrsrpsgsisylpsfftkl--estspmvkskqefrklssggg 402
QY 2301 EPEYNISOPGTIFNMPLTGTGLTYRSGAOVHASTNMGLEATIRKALMGKYPQWEE- 2359
Db 403 dsdmaaaqptgfplpavttsvgsrshsfadpas--nlgledlirkalmgsfdkved 461
QY 2360 -----SPPLSANAFNPLNASALPAMPITAADGRSDHTLSPGGG--GKAKVSGRPSRRK 2413
Db 462 hgvvmshpv--gimpgaastav-----vtsearredegepsphagvckpklknsrsk 513
QY 2414 AKSPAPGLA--SGDRPPSVSVHSGDCNRRPTLNRVWEDRPPSSAGSTPPYPLNRL 2471
Db 514 skspplgsgylgterpsvsvshsgdyhrqtp--gwawedrpstgtqfypnitirm 571
QY 2472 QAGVMASPPPPGLPAGSGPL--AGPH--HAWDEPKPLCSQYETLSDSE 2517
Db 572 ----lssptqlacapsaitqaqhqnriwerrepapllsaqyetlsdsd 618

RESULT 6
AAB42491
ID AAB42491 standard; Protein; 3266 AA.
XX AC AAB42491;
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF2255 polypeptide sequence SEQ ID NO:4510.
XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX KW vulnary; antiposrotic; antiparkinsonian; neurotropic; neuroprotective;
XX KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
XX KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
XX KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
XX KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX KW cardiovascular disease; diabetes mellitus; erythematosis; SCID; AIDS;
XX KW cholesterol ester storage; systemic lupus erythematosus; infection;
XX KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX KW thrombosis; contraceptive.
```

Db 942 kqadrfdvspnsii-----krdslrksvrdl--epgevpsdsdedgkshspsrasal 995
Qy 323 -----RAKESKREYVEKOPPEIRKORLOQWOSRVQSGSL-----SMSAAR 367
Db 996 yessrlsflldredkrlenderlssierknkfysfaldkitptdkalleraksissr 1055
Qy 368 SEHVESEIIDLSE-----OENLEKOMROLAVIPPMYLDADQOIRIKFINMGLADPMKVY 423
Db 1056 een--wslfdwsrfanfnnkdkvdsapripwmywmkkkkr--tdsegkmdkdkdh 1112
Qy 424 KDRQVNMWMSOEKTFREKFMHPKFNGLTASFLERTVAECVLYLYLTKNENYKSLV 483
Db 1113 keee-----qergelfasrfl-hsifeqskrlq-----hlerkeedsdfls 1154
Qy 484 RRSYRRRKSQQQQQQQQQQQQQQQQOQP-----MPRSSQBEKDEKEKEKEAEE 533
Db 1155 griy--gk-----qtsegansttdsileqpvlfhsrfmeltmqgkkekdkpkeveq 1208
Qy 534 EKPEVENDKEDLLKEKTDGSDNDENDEKAVASKORKTANSOGRRKGRITRSMANEANSE 593
Db 1209 ---dten-----hpktpesapenk-----selktp 1231
Qy 594 EAITPQQAELASLMELNESSRWTEEMETAKKGLLEHGRNMSAIARMVGSKTIVSOCKNFY 653
Db 1232 psvgpvsv-----tvvtlesap-----salektgdktv----- 1260
Qy 654 FNYKKRONDELTOOHKLKMEKERNARRKKKAPAAASEEAAFPVVEDEMEASGVSGN 713
Db 1261 -----eaplvteektvepatvseeakpasepapa 1289
Qy 714 EEMVVEEAELHALHAGNEVPRGECGPATVNNSSDTPESPHTEAAKTGQNGPKPPATL 773
Db 1290 pveqleqv-----dippga-----dpckeaaamp-----agveegsgdqp--yl 1328
Qy 774 GADGPPGPPPTPPRTSRAP--IEPTPASEATGATPPAPPPSPAPPPVPVPEKEKEEET 831
Db 1329 dak-----pptpgasfsgaenvdpepds-----tqlskpaqkseaneakpdpdat 1377
Qy 832 AAAPP-----VEEGEOKPPAAAEELAVD---TGKAEEPVKSECTEEAEPEGKAKGDAA 883
Db 1378 adaepdankaeapessppasedlevdpvvaakdkkpnkkskrktpvqaaavivekpv 1437
Qy 884 EATBEGALKAEKKEGGGRATTAKSSGAPQSDS--SATCSADEVDEAEAGGKNRLLSPR 942
Db 1438 trkseridreklnrsprgeaqllelmeaekitrtasknsaadle-----hpep 1489
Qy 943 SL-LTPT-----GPRANASQK-PLD-----LKOLKORAAAIPIQVTK 980
Db 1490 slpstrrrnrvsvyatmgd--henrsvpkepvqevprvtrkrlerelgeaaavp--ttpr 1546
Qy 981 VHEPPR-----BDAAPTTPAPPAPP-----PQNL-----QPESDAP 1012
Db 1547 rgrpkrtrradeeeneakepaetlkpegwrsprsgktaaggpggqgkknepkvdat 1606
Qy 1013 QOPGSSPRG-----KSRSPAPPADKFAFAEAQKLPDPPCWTGSLPFPVPPPREVIKASP 1067
Db 1607 rpeattevpgqigvkessmepkaaeaeagseqkdrkd--agtdknnpbetapvevvekkp 1664
Qy 1068 HAPDPSAFSYAPPHPPLGLLHDTPARVLP RPPTTSNPPPLISSAKHFSVLERQIGALISO 1127
Db 1665 -apeknks-----krgsrnrsrlavdkksa1kn-----vdaavsprgaaagageres 1711
Qy 1128 GMSVOLHVPYSEHAKAPGVPTMGLPLMDP-----KKLAPPSG-----VKOEOL-- 1172
Db 1712 gy-vavspeksespqkedglsqklsdvpdpokekedvsaasgspseatqlakgmeleq 1770
Qy 1173 -----SPRGAGGPPELSGVPTAQEAASVLRGTALGSV----- 1203
Db 1771 avehlaklaeasasaaykadapegla--pedrpkpahasetelaaalgsindisgepe 1828
Qy 1204 -----PGGSITKIPSTRV-----PSDSAITYRGSTTHGTPADVLVYKGTI-TRIIGED 1250
Db 1829 nfpappppypgesqldlqpagaqalqpse-----egmetdeavsvgileteaates 1878

Qy 1251 S-----PSRLDRGRDSDSLPKGHVIEYEGKKGHVLSYEGGMSVTQCCKEDGRSS 1298
Db 1879 srppvnapdpagpdtckeargnsetshtsvpeakgsk-----eventlvrkdkgrq-- 1930
Qy 1299 GPPHETAAKRTYDMMEGRVGRAISSASIEGLMGRAIPPERHSPHHLKEQHHRG---SI 1355
Db 1931 -----ktrsrkrntnkkv-----vapveshvp-----esndaqagespaa 1966
Qy 1356 TOGIPRSYVEAQEDYLREAKLLKREGTPPPPPSRDLTEAYKTOALGPLKLPKPAHEGLV 1415
Db 1967 negttvqhpapeq-----eekqsexphstppscdsckskipste----- 2006
Qy 1416 ATVKEAGRSIIEHPREELRHT-----PELPLAPRLKEGSITQGTPLKYDYDTGASTGSRK 1470
Db 2007 -----nssqeiseveertptkasvpdpdlppppqp-----apvdeepqa-----r 2044
Qy 1471 HDVRSLSIGPGRTPP--PVHPLDVMDARALERACYEESLAKSRPGTASSGSGSIARGAPV 1528
Db 2045 frvhsliesdpvtppsdpbsipptlpsv-----taaklsppvasg-- 2084
Qy 1529 IVPGLGKPROSP-----LTVEDHGAPFAGHLPRGSPVTMREPTPRLQEGSLSS--SK 1578
Db 2085 -----gibqspptkvwtwitrqe-----epraqstpspalppdtkasdvds 2127
Qy 1579 ASQDKRLSTPREIAKS-----PHSTVPEHHPHPISPYEHLRLGVSGVDLYR 1625
Db 2128 sstllrkiilmppkyvaatsvtstsvttaeaepvsaapclheappppvd-----skkplee 2181
Qy 1626 SHIPLAFDPTSIPIRGIPLDAAAAYYLPRLAPNPYPHYLPYLYRGYPDPTAALENQRTI 1685
Db 2182 ktappvtunseiqasevivaadkekvaplapkit-----svismrmpvsidlensski 2234
Qy 1686 INDYITSQOMHNTATAMAQRAQDMLRGLSPRESSLALNVAAGPRGLIDLSQVPHLPVLVP 1745
Db 2235 -----tlakpaptltgl-----vsaltglvnlsvlp--vnalkg 2267
Qy 1746 PTPGPTATAMDRLAYLPTAQPFSSRRHSSPLSPGPGTHLTPTTTTSSSERERDRDRD 1805
Db 2268 pvkgsvtllkslvs-----tpagpvnlkgpv----- 2294
Qy 1806 RDREKESILTSTTVEHAPI-----WRPGTQSSGS-----SGSS 1841
Db 2295 -----nvltpgvnlvtppvnaatgtvnaaogtvnaaasavnatasavtvtagavtaas 2347
Qy 1842 GGGGS-----SSRPASHSHAHQHSPISPRTQDALQORPSVLHNTGMKIIT 1888
Db 2348 ggvattgtvtmagaviapstckckgrasanensrflhpgsmpvldirpa---dag--sagql 2403
Qy 1889 AVEPSKPTVLRSTS---TSSSPVRPAA---TFPPATHCPLGGTLDGVYPTLMPEVLLPKEA 1942
Db 2404 rvntsegvllsysgqktegpgqrisakisqlppas-----amdliefqgsqsvskvpkds 2457
Qy 1943 PRVARP--ERRPADTGHAPLA-----KPPARSGLPEAPASSPSKSGSEPRPL---- 1984
Db 2458 vtasqpskpgqpagyannvathstlvtltaqynaspsvissvk--adrpsal--ekpepilis 2515
Qy 1985 -----VPPVSGHATIAITPA----- 1999
Db 2516 vstpytqggtkvltcggintppvlvhnqlvltpsvlttnkkladpvtlkietkvlqpanl 2575
Qy 2000 -KNLAPHUASPPPPAPASADP--HREKTQSKPF--SIOELELSLGVHSGSYSPGEGVE 2054
Db 2576 gstlcpvh-----ppalpsklptevnhvpsgspisadrtvshlaaakldahsprpsgpgps 2631
Qy 2055 PVSPVSPSLTHDKGL-----PKH--LEBLDKSHL- 2082
Db 2632 sfrashpsstastalstnatvmlaagipvpqfissihpeqsvimpphsaitqtvsishls 2691
Qy 2083 EGELRPKOPG-----PVKLGGEAAHPLPHLPPLPESQSSPPLLOTAPGVKGHORVVTLAQH 2138
Db 2692 qgevrmntptlpsitcitysirpealhspr-apl---qp-----qqievra--- 2730

QY 2139 ISEVITQDTRHHPOQLGAPLAPLYSFGASCPLVLDLRRPPSD---LYLPDPDHGAPAR 2195
Db 2731 -----pqrastpqap-----agypalasqhpeeeerhyhlpvaratapvq 2771
QY 2196 GS-----PH-----SEGKRSPENKTSVLGGEDG 2221
Db 2772 sevlvmqseyrlhpytvpdrvm/hphvtavseqpraadgvykvpaskap-----qqpg 2827
QY 2222 IEPVSPPEGMTEPGHRSNAV-----YPLLRYDGEQTEPSRMGSKSPGNTSQ-----PP 2269
Db 2828 keaaktadakaaptcpapvpvvpplpapapaphge-----ariltvpsnqlgpltp 2883
QY 2270 AFFSKLTESNSAMVSKQKQIEINKLNTNHRNE--PEYNISQPGTEIFNMPALTGTGLMTY 2327
Db 2884 -----vvvthgvqi-----vhsqgelfeyrygd-----irty 2911
QY 2328 RSQAVOEH-----ASTNMGLEAIIRKMGKYDQWEE-SPPLSANAFNPLNASLPAAMP 2382
Db 2912 hpaqlthtqfaasvlglsrktkaagpppegeplqppgvqstqapapppcpssq- 2970
QY 2383 ITAADGRSDHTLTSPGGGKAK---VSGRPSSRKAKSPAPGLASG--DRPP-----SVSSV 2433
Db 2971 -----lgpgqppsskmpvsgaekgtgtgveqprlpagpanrpphtqvqra 3019
QY 2434 HSE-GDCNRRTPLTNRVWEDRPSSAGSTPFYPYNPLIMRLQAGVMASTPPPGLPAGSGPLA 2492
Db 3020 qaeetgtsfsvpsvsmkplvslptqtapqplfvpttsq-----patppg---lv 3069
QY 2493 GPHAWDEEPK-----PLICSQ 2509
Db 3070 lphtefqapqkqdsphitsq 3090

RESULT 7
AAB50362
ID AAB50362 standard; protein; 3118 AA.
XX AC AAB50362;
XX DE 12-MAR-2001 (first entry)
XX DE Human SRCAP.
XX KW Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;
KW CBP regulatory element; CREB binding protein; CBP; ATPase;
KW transcription activation; DEAD box RNA dependent helicase;
KW adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.
XX OS Homo sapiens.
XX PN WO200073467-A1.
XX PD 07-DEC-2000.
XX PF 25-MAY-2000; 2000WO-US14719.
XX PR 27-MAY-1999; 99US-0136620.
XX PR 25-MAY-2000; 2000US-0579181.
XX PA (UYSL-) UNIV SAINT LOUIS.
XX PI Chirivia J, Yaciuk P;
XX DR WPI; 2001-061545/07.
XX DR N-PSDB; AAC89859.
XX PT Snf2 related CAMP regulatory element (CREB) binding protein (CBP)
PT activator protein, capable of co-activating CREB binding protein,
PT useful for modulating transcription and for affecting viral infection -
XX Claim 6; Page 77-86; 103pp; English.
XX CC The present sequence is an Snf2 related CREB (cAMP regulatory element)

CC binding protein (CBP) activator protein (SRCAP) polypeptide. It has
CC ATPase activity and is capable of activating transcription. SRCAP
CC polypeptides are useful for activating transcription in a cell, for
CC enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated
CC activation of transcription in a cell, for treating a patient having a
CC disease involving a function such as insufficient transcription of a
CC gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent
CC helicase, adenoviral DBP protein, beta-actin or a nuclear receptor
CC affected by SRCAP protein. Compounds that modulate SRCAP function, such
CC as antibodies, antisense molecules, polynucleotides or ribozymes, are
CC useful for treating diseases mediated by SRCAP-activated transcription,
CC for example, infection by adenovirus, hepatitis C virus, human
CC immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or
CC hepatitis B virus.
XX Sequence 3118 AA;
SQ

Query Match 3.6%; Score 477.5; DB 22; Length 3118;
Best Local Similarity 18.6%; Pred. No. 4.8e-18;
Matches 597; Conservative 330; Mismatches 1109; Indels 1169; Gaps 136;

QY 137 KDRSLTGKLEVPSPPPH--TDP-----ELELVPPRLSKEEL-IQNMDRVREI 183
Db 2 qdssldg-----ppgpdgatvplgflsqaadlankgpkwekshaeiaekhaei 55
QY 184 TMVEQOISKLKKKQOOLEEAAKPEPEKP-----VSPPIESKH-----R 224
Db 56 ---etriaelrkegfswlsrlpkvpeprpkghwdylceemqwisadfaqerwrgvar 112
QY 225 SLVQII---YDENRKAEEAHRILEGLGPQVELPLYNQSDTRYHENIKINQAMRKLI 281
Db 113 kvrmvirhheeqrkeearreeqa---kiriastmakdvrgfwsnv-----ekv 161
QY 282 LYFKRRNARKQWKQFCORYDQMLEAKKVERTENPRRAKESKY----- 329
Db 162 vqfkqsrleekrkaldhldfivgqtkeysdlisqnlngltsskagsgpcilgssaa 221
QY 330 -----REYKEQFPEIRKORE-----L 346
Db 222 sspppasrlddedgdfqpddeeddeetieveeqgndaeagrrreillrreglpl 281
QY 347 QERMQSRVQRGSG-LSMSAARSEHEVSEIIDLGEQENLEKQMRQ-LAVIPMLYDADQ 404
Db 282 eellrslpqllegpspsqpsshd-sdtrdg--pegaeepqvleikppp--savl 336
QY 405 QRIKFNNGLMADPMKYKDRQVNMWMSQEKETFRKFMQHPKFNGLIASFLERKIVA 464
Db 337 qrnk-----qpwhpdeddefta-neeaeadeetiaaeqlegdevdhamsela 386
QY 465 -----ECVLYYL-----TKKNENYKSLVRSYRRRGKSOQQOQQOQQOQQOQQO 509
Db 387 regelsmeellqqayagayagsgssededevdandsdcepegvveeepqedsqs 446
QY 510 QMPRSSQEEKDEKEKEA-----EKEEEKPVEVNDKEDLLKEKTDTSQ-----E 556
Db 447 dsvedrsedeelheeeetsgssaseeseeesadaqsqsqadeeeddfvgyellar 506
QY 557 DNDEKEAVASGRKT-----ANSQGRKG----- 580
Db 507 deeqseadagsgppptgttligpkkeidtaaaeslqpktylattgvtktpillrgq 566
QY 581 -----RITRSMANEANSSEAI-TPOOSAELA-----S 606
Db 567 lreyqhigldvltmyekkingiladenglgtigtisllanlacekgnwgphtliivpts 626
QY 607 MELN---ESSRW-----TEEMETAKKG-----LLEH---G 631
Db 627 vmlnwemelmkwcpsfkiltygaqkerklrkgwtkpnafhvcitsxlvldhqafrr 686
QY 632 RNWSAIARMVGSKTVSQCKNF-----YFNKKR-----QN-----L 662
Db 687 knw-----ryllldeaqniknfksgrwsgallnfnsgrillltgtplqnslmelwslmhl 742

QY 1235 ADVLYKGTITRIIGEDSPRLDRGREDSLPKGHVIV-----EGKKGHVLSYEGGMSVTQCS 1290
Db 1317 aasaaqalatlapmaaaqatallapaspapplaplpvlapspgaapavlassqtpvpmvmaps 1376
QY 1291 KEDGRS---SSGP---PHETAAPKRYTDMMEGRVGRAI--SSASIEGL-MGRAIIPPERHSPH 1343
Db 1377 stpgtstlasaspvaptvlpavssqtmlpavpvpplspastqtalalalapt-----1431
QY 1344 HLKEQHHRGISTQIGIPRSYVEAQEDYLREAKLLKRGCTPPPPPSRDLFEAYKTQALG 1403
Db 1432 -----199s-----spqstslgtnpgqpfptqtls-----1458
QY 1404 PLKLPAPHEGLVATYKEAGRSITHEPRELRTPELPLAPR--PLKEGSTTQGTPLKYDTG 1462
Db 1459 ---lcpa-sslvpt---paqtislapgpplqptqtislapapplapaspavgpapahtitl 1511
QY 1463 ASTTSGSK-----HDVRSLIGSGRTFPVPHPLDVMDARALERACYESELKSRGTASS 1518
Db 1512 apassasallapasvqtitlspa-----pvptlqpaaqtlalalapasqtspasqasl 1564
QY 1519 GGSIRGAPVIV-----PELCKPRQSPLTYEDHGAPFAGHLPRGSPVTRREPT 1566
Db 1565 vvasgaaplpvmvmsrlpvsksdepdtitlrgspsspststatsfggprprqp-----ppp 1620
QY 1567 PR-----LQF-----GSLSSSKASQDRKLTSTPREIAK--SPH 1597
Db 1621 prspfyldsleekrrqrserlerifqlseahgalapvygtevldftclpqpvaspiqr 1680
QY 1598 STVPEH-----HPHPTSPYHLLRGVSGVDLYKSHIPLAFDPTSPRGIFPLDAAA 1647
Db 1681 spgshptfwtteahravlfpqrdqlseilfierfivmp-----pveapppslh 1732
QY 1648 AYLPRLHAP-----NPTYPHLYPPYLI-----RGYPD-----TAAL 1680
Db 1733 achppwlapraqafqeqlaselwpraplhrivcnmtqfplrligydcqlqlav 1792
QY 1681 NRQ-----TIINDYITSQ-----QMHNTATAMAQADMLRG-----1712
Db 1793 lqqlkaeghrvltfqtmrldvleqfityghlylridgstveqrqalmerfnadkri 1852
QY 1713 ----LSPRESSLANYAGPRGIIDLSQVPHLPVLVPTTGPATAMDR-----1757
Db 1853 fcfilistrsgvgvnlgtadvfydsdw-----nptmdaqaqdrchrigqtrdvh 1903
QY 1758 -----LAVLPTA-----PQPFSSRRSSS- 1775
Db 1904 lyrliserteveenilkkanqkrmldmaiegnftayfkqqtirelfdmpleepsssv 1963
QY 1776 PLSP-----GGPHTLTKTPTTSSSERDR-----1800
Db 1964 psapeeeetvaskthileqalcradeediraatqakaeqvaelaefnendgfpag 2023
QY 1801 -----DRDRDREREKSLTSTTVEH-----1823
Db 2024 eeagrpgaedeemsraeqeiaalveqlptieryamkfleasleevsrealkqaeeqvea 2083
QY 1824 -----APTWR-PGTQSGSSG--SSGGGGSSSPASHAHQH--SPISPTQDA 1870
Db 2084 rkdldakeevfirlpqeeegpgagdesccgtggthrrskakaperpgtrvserlrga 2143
QY 1871 LQQRPSVLHNTGMKGIITAVPSKPTVLKSTST-----SPVPRPAATFPATHCPLGLGTLDG 1927
Db 2144 raetogahnp-----visahq-----trsttprcpaervrpraprptpasapa 2193
QY 1928 VYPTLMPEVLLPKEAPRVARPERPRADTGHAFKAPPARSGLEPASPSKSGEPRLVPP 1987
Db 2194 alpal---vpvpsavpvisanpntilpvhilpspppsqipccspa--ctpppactp 2248
QY 1988 VSGHATIAITPAKNLAPHASP---DPPAPPASAS-----DPHREKTSQKPFISIEL 2036
Db 2249 ppahtp---ppaqtclvtppspillgppsvpissavtnlplglrpeaelcaqalaspsel 2305

QY 2037 ELRSLGYHGSS-----YSPGEGVPVS-----PV-----SSPSLTHDKG-LPKHLEELD 2078
Db 2306 elasvasasetstslvppkdlipvaveilpvseknsltpsaapstleagslngqeqea 2365
QY 2079 KSHLGEELRPKQGPVKVLGGEEAHLIP-----HLRPLPESQSPSSSLQLQTAFCVKGHQRV 2132
Db 2366 pdaaegtltitlv-----egeelplcvsesnglielpsaasdeplqpleadr-----2413
QY 2133 VTLAQHIEFVITQDTRHHPOOL-SAPLPAPLYSPGASCPLDLRRPPSDLYLPPPDHG 2191
Db 2414 --tseelteaktptsspekpqelvtcaevastssats-----spep 2455
QY 2192 APARGSPHSEGRKRPENKTSV-----LGGEGDIEVPSPPEGMTPEGHSRSAYVPLL 2245
Db 2456 spar-----pprrrtsadveirgqgrpg-qppgpkvrlkplgrlvtvv-----2499
QY 2246 YDGQTTPSRMGSKSPGNTSQ--PPAFFSKLTESNSAMVSKK-QEINKKLLTHNNRPE 2303
Db 2500 ----eekelvrrrrqrgaastlvpgvsetsaspgsvrsmgspessppiggpceaaaps 2555
QY 2304 YNISOPGTEIFNMP-----AITGTG-----LMTYRSQAVQEH-----ASTN 2339
Db 2556 ssipppqgqfiarhielgvctggspengdallaitpavkrrrrgrppkknrspadag 2615
QY 2340 MGLEAIIRKALMGKYDQWESP-PLSANAFNPLNASASLPAAMPITAADGRSDHTITSPG 2398
Db 2616 rgvdeapsstlkgktngadpvgpetlivadvleqlipgqplgqpvhprnpillsp- 2674
QY 2399 GGGKAKVSGRPSRRAKS--PAPGLASGDRPPSVSVHSEGCNRRPTLTNRVWEDRPSSA 2457
Db 2675 --vekrirgrpp--kardlipg-----tissa-gdgnsestr-----qppphps 2714
QY 2458 GSTPPFPYNPLMLRLOAGV-----MASPP-----PPGLPAGSGP 2490
Db 2715 pltlp--pllivcatvantvtststpsbkrkrgrppknppspr 2760

RESULT 10
AAB50363
ID AAB50363 standard; protein; 2972 AA.
XX
AC AAB50363;
XX
DT 12-MAR-2001 (first entry)
XX
DE Human SRCAP.
XX
KW Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;
KW CAMP regulatory element; CREB binding protein; CBP; ATPase;
KW transcription activation; DEAD box RNA dependent helicase;
XX adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.
OS Homo sapiens.
XX
XX WO200073467-A1.
XX
PD 07-DEC-2000.
XX
PF 25-MAY-2000; 2000WO-US14719.
XX
PR 27-MAY-1999; 99US-0136620.
PR 25-MAY-2000; 2000US-0579181.
XX
XX (UYSL-) UNIV SAINT LOUIS.
XX
PI Chrivia J, Yaciuk P;
XX
DR WPI; 2001-061545/07.
DR N-PSDB; AAC89860.
XX
XX Snf2 related camp regulatory element (CREB) binding protein (CBP)
PT activator protein, capable of co-activating CREB binding protein,
PT useful for modulating transcription and for affecting viral infection -

QY 1530 VPELGKPROSLTYEDHGAFAGHLPRGSPVTMR--PTPRIQEGSLSS--KASQDRK 1584
 Db 1040 mprvrkpktp-----tprkmtstmpelnptisraeamlqtrpnpqpsk 1086
 QY 1595 LT-----SPTREIAKSPHSTVPEHHPH-----PISPYEHLRLRGVGVDLIYS 1626
 Db 1087 lvevnpksedaggaegpghmlr-phvfmpevtmdydlprvpqugiin----- 1136
 QY 1627 HIPLAFDPTSPRGIPLD 1644
 Db 1137 --pmslsetnioncgkpv 1152

RESULT 13
 AAB60568
 ID AAB60568 standard; Protein; 1404 AA.
 XX
 AC AAB60568;
 XX
 DT 27-APR-2001 (first entry)
 XX
 DE Human megakaryocyte stimulating factor (MSF, CACP).
 XX
 KW Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;
 KW MSF; megakaryocyte stimulating factor; synovial lubricant;
 KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
 KW antiarthritic.
 XX
 KW Homo sapiens.
 OS
 XX WO200107068-A1.
 FN
 XX 01-FEB-2001.
 PD
 XX 21-JUL-2000; 2000WO-US20002.
 FF
 XX 23-JUL-1999; 99US-01453328.
 PR
 XX 19-JUL-2000; 2000US-01453328.
 PA (UYCA-) UNIV CASE WESTERN RESERVE.
 XX
 XX Warman ML;
 PI
 XX WPI; 2001-182721/18.
 DR
 XX
 XX
 XX New composition comprising the camptodactyly-arthropathy-coxa
 PT vara-pericarditis protein in combination with an anesthetic, useful for
 PT treating osteoarthritis, or as lubricants of tissue and joints -
 XX
 XX Example 1; Page -; 34pp; English.
 PS
 XX
 XX The invention relates to a method of treating osteoarthritis via the
 CC administration of a composition comprising the camptodactyly-arthropathy-
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 CC The composition may further comprise a local anesthetic. The composition
 CC of the invention may be administered via intra-articular or intravenous
 CC injection. The human CACP protein is identified in the invention as
 CC being megakaryocyte stimulating factor (MSF). The gene encoding
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
 CC this gene are responsible for the heritable disorder camptodactyly-
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)
 CC acts as a synovium lubricant, and can be used to lubricate tissue and
 CC joints in the treatment of osteoarthritis. The composition may be
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
 CC loss of range of movement or joint damage). The present sequence
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).
 CC Note: This sequence is not given in its entirety in figure 4 of the
 CC specification, although a GenBank accession number was given. This
 CC sequence was therefore obtained from GenBank (U70316).
 XX
 XX Sequence 1404 AA;
 SQ

Query Match 3.2% Score 428 DB 22 Length 1404;
 Best Local Similarity 20.8% Pred. No. 1e-15;
 Matches 266; Conservative 136; Mismatches 516; Indels 360; Gaps 54;

QY 476 NENTKSLVRRSYRRGRKSO-----QQQQQQQQQQQQQQQQPPRSPQSQEKDEKE 524
 Db 126 sgtikstkrspkpnkktkkvieseeiteehsvsenqsssssssstiwwkks 185
 QY 525 KEKEAEKE-EKEPEVENDKEDLLKEK-----TDDT-SGENDKEKAVASGKRKTANSQ 575
 Db 186 knsaanrelqtklkvdknkrtkkktpkppvdeagldngdfkvttdsttghnk 245
 QY 576 GRRKGRITRSMANEANSEAITP-QQSAELASMELESNRWTEEMETAKGLLEHGNW 634
 Db 246 vstspklt--takpinrpslppnsdtsktsitvknkettvtkttnkqtdgkek 303
 QY 635 SAIAVMGSKTVSQCKNFYFNYKKRQNLDEILQOHKLME---KERNARRKKKAPAAAS 691
 Db 304 ttsaketgsiekttsakdl-----aptskvlaktpkaetttkgpaltpkepttpk 356
 QY 692 EEAAPPPVVEDEMEASGVSGNEEMVEEAALHAGNEVPRGECGSPATVNNSDTESI 751
 Db 357 epasttp-----kept-----pttiksaptpke 380
 QY 752 PSPHTEAAKDTGONGKPPATLGADGPPPGPTPPRRTSRAPIEPTPASEATGAPTP-- 809
 Db 381 paptttsaptpkepaptp-----kepaptpkepaptppttkpaptpct--tsaptpke 434
 QY 810 PAPPSPSAPPVVPVPEKEKEETAAAPVVEEGEQKPPAAEELAVDTGKAEPPVKSECTEE 869
 Db 435 paptpkpkpaptpke--paptpkepttpkepaptpkepaptpkepaptpkepaptp 485
 QY 870 AEEGPAGKDAEABAEATGALKAEKKEGGSGRATTAKSSGAPQSDSSATCSADEVDEA 929
 Db 486 apkxpaptpkepaptpkepaptp-----pspptkepaptppttksaptpke----- 537
 QY 930 EGGDKNLLSPRPSLLPTGPPRANASPOKPLDLKQLKRAAAIPPIQVTKVHEP---P 985
 Db 538 -----appttksaptpke-----spttkkepaptp 565
 QY 986 REDA--APTTPAPPAPPQNLOPESDAPQQSGSPRKSRSAPPADKFAFAEAQKLP 1043
 Db 566 kepaptpkpkpaptpkepaptpkepaptpktpkpkpaptpkepaptpkpklt 625
 QY 1044 GDPP-----CWTSGLPFPVPPPREVIKASPHADPSAFSAPPCHPLPLGLHDTARVLP 1099
 Db 626 pttbeklaptpkepaptpkepaptp-----peepaptp-----tpkaapapntpk 679
 QY 1100 PTISNP-PPLISSAKHPSVLERQIGALSQMSVOLHVPYSEHAKAP-----VGPVTMGLP 1153
 Db 680 paptpkepaptpkpa-----pttketaptpkgtapttkp 720
 QY 1154 LPMDPKKILAPPSGVKQEQLSPRGQAGPPESLG---VPTAQEASVLRGTALGVSPGGSITK 1210
 Db 721 apttpkpa-----kelaptttkptstsdpaptp-----kga----- 758
 QY 1211 GIPTRVPSDAIYRSGITHGTADVLYKGTITRIGEDSPSLDRGREDSLPKGHVIY 1270
 Db 759 --ptt--pkpaptp-----tpkepaptpkgtapttkkepaptpkpkpapkepap 804
 QY 1271 EGKKGHVLVEGGMSVTCQSKEDGRSSGPPHETA--APKRTYDMMEGRVGRASSASIE 1328
 Db 805 -----tttkgpcstt-----sdkpaptpkecaptpkepaptpkpkpaptpctp-- 850
 QY 1329 GLMGRAIPP---ERHSPHLLKEQHIRGSITQIGIPRSVVEAQEDYLREAKLLKEGTP- 1384
 Db 851 -----pptsevstpttkpttkhkspsdestpe-----lsaeptpk 887
 QY 1385 ----PPPPPSDLTEAYKTQALGPLKPAHEGLVAVVKEAGRSIHETPRELHPTPE-L 1439
 Db 888 alenspkpvgvpttkt-----paatkpe---mttakd-----kterdlrttptt 931

QY 1440 PLAPRLKESGIT-----QGTPLKYDTGAS-----TTGSKK 1470
 Db 932 taapkmktetatttktattgtvttdttdtpfittkitttlapkvtttkk 991
 QY 1471 HDVRLSGPRTFPVHPLDMADARALERACYEESLKSRPGTASSGGS-IARGAPVI 1529
 Db 992 ittteimkpkpeet-----akpkdratskattkpkqpkapktstskpkk 1039
 QY 1530 VPELGKPRQSPLTYEDHGAPAGHLPRGSPVTMRE--PTPLQEGSLSS-----KASQDRK 1584
 Db 1040 mprvrkpkttf-----tprkmtstmpelntpsriaeamlqttrpndtpnsk 1086
 QY 1585 LT-----STPRIAKSPHSTVPEHHHP-----PISPYEHLRLRGVGVLDLYRS 1626
 Db 1087 lvevnpksedaggaetpnhllr-phvfmpvtdmdylprvpnqgliin----- 1136
 QY 1627 HIPLAFDPTSTPRGIPLD 1644
 Db 1137 --pmlsdetnfcngkpv 1152

RESULT 14

AAB29773
 ID AAB29773 standard; Protein; 1404 AA.
 AC AAB29773;
 XX 28-FEB-2001 (first entry)
 XX Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.
 KW Human MSF; megakaryocyte stimulating factor; tribonectin;
 KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
 KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
 KW friction coefficient reduction; gene therapy; antiarthritic;
 KW osteopathic.
 OS Homo sapiens.
 XX WO200064930-A2.
 PN 02-NOV-2000.
 PD 24-APR-2000; 2000WO-US10953.
 PF 23-APR-1999; 99US-0298970.
 PR (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
 PA Jay GD;
 PI WPI: 2001-024673/03.
 DR N-PSDB; RAC81498.
 XX Novel tribonectin polypeptide useful as lubricant for treating
 PT osteoarthritis, comprises O-linked lubricating moiety
 XX Claim 3; Page 7; 47pp; English.
 PS The invention relates to a human tribonectin which is a product of
 CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
 CC gene. The tribonectin has at least one O-linked oligosaccharide
 CC lubricating moiety and has a polypeptide sequence comprising 1-76
 CC repeats of a motif having at least 50% identity to the sequence KEPAPT
 CC (AAB29774). The invention also relates to a nucleic acid encoding a
 CC human MSF-derived tribonectin; a biocompatible composition comprising a
 CC human tribonectin for inhibiting tissue adhesion formation; and a method
 CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
 CC measuring the amount of MSF or its fragment in a biological sample of a
 CC mammal, wherein an increased amount of MSF compared to a control
 CC indicates the presence of or predisposition to developing
 CC osteoarthritis. The tribonectin and DNA encoding it are useful in the

CC treatment of osteoarthritis, where they may be used for lubricating
 CC mammalian joints, such as articulating joints of humans, dogs or horses.
 CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is
 CC useful for inhibiting adhesion between two surfaces such as the injured
 CC tissues of a mammal, where the injury is caused by a surgical insertion
 CC or trauma, or an artificial device e.g., an orthopaedic implant. In
 CC particular, one of the surfaces is pericardial tissue. DNA encoding a
 CC tribonectin may be used in gene therapy. The present sequence represents
 CC human MSF.

XX Sequence 1404 AA;

Query Match 3.2%; Score 428; DB 22; Length 1404;
 Best Local Similarity 20.8%; Pred. No. 1e-15;
 Matches 266; Conservative 136; Mismatches 516; Indels 360; Gaps 54;

QY 476 NENYKSLVRSYRRRKSQ-----QQQQQQQQQQQQQQQQPPRRSQEKEDEKE 524
 Db 126 sqtksttkrskppnkkkkvieseeiteehsvsenqssssssssstlwkikas 185
 QY 525 KEKEAEKE-EKPEVENDKEDLLKEK-----TDDT-SGEDNDEKAEVASKRKTANSQ 575
 Db 186 knsaanrelqkklkvdknkrkkktpkpvpvdeagsgldngdfkvttdtsttqhknk 245
 QY 576 GRRKGRITRSMANEANSEEAITP-QQSAELASGMELESRSRWTEEMETAKKGLLEHGRNW 634
 Db 246 vstspklt--takpinrpslppnsdtsktsltvnkettvetkttntktsdgtkek 303
 QY 635 SAIAWMVGKTVSQCKNFYFNYKKRONLDEILQOHLKME---KERNARRKKKAPAAAS 691
 Db 304 tsketqtsiektakdl-----aptskvlaktpkaettkgpalttptkptk 356
 QY 692 BEAAPPVVEDEMEASGVSGNEEWEAEALHAGNEVPREGCGATVNNSSDTSI 751
 Db 337 epasttp-----kept-----pttkksapttkpe 380
 QY 752 PSPHTFAAKDTGQNGPKPPATLGADGPPPPPTPRRTSRAPTEPTPASEATGAPTTP-- 809
 Db 381 paptttkksapttkpeapttt-----kepapttkpeapttkpeaptt- tksapttkpe 434
 QY 810 PAPPSPPSAPPVVPVKEKEEETAAAPVVEEGEEKPPAAEEALAVDTGKAEVPKSECTEE 869
 Db 435 paptttkkpkapttkpe--papttkpepttkpeapttkpeapttkpeaptt----- 485
 QY 870 AEEGPAKGDAAEAATAEGALKAEKKEGSGRATTAKESSGAPQSDSSATCSADEVDEA 929
 Db 486 apkkpapttkpeapttkpeapttkpe-----pspttkpeapttkksapttkpe----- 537
 QY 930 EGGDKNRLISPRPSLLTPTGDPPRANASPKPLDLKQLKORAAAIPPIQVTKVHEP----- 985
 Db 538 -----aptttkksapttkpe-----aptttkpeapttkp 565
 QY 986 REDA--APTTPAPAPPQQNLQPSDAPQQPGSSPRKSRSPAPPADEKFAFAEAQKLP 1043
 Db 566 kepapttkkpkapttkpeapttkpeapttkkpkaptkpkapttkpapttkpkklt 625
 QY 1044 GPPP---CWTSGLPFPVPPREVIVKASPHADPPSAFSAFYPGHPPLGLHDFARVLP 1099
 Db 626 pttpeklapttkpeapttkpeelapttpeeptptt-----peepapt-tpkaaapttkpe 679
 QY 1100 PTISNP-PPLISSAKHPVLERIGAIQSGMSVQLHVPYSEHAKAP-----VGPVTMGLP 1153
 Db 680 paptttkpeapttkpepa-----pttpketapttkpghtapttkp 720
 QY 1154 LPMDPKKLAPSGVKQEQQLSPRGQAGPPESLG---VPTAQEASVLRGALGVSPGSGITK 1210
 Db 721 apttkpkpap-----kelapttktsdskpaptt-----kgta----- 758
 QY 1211 GIPSTRVPSDAITVRSITHTGTPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVYI 1270
 Db 759 --ptt--pkpapt-----tpkepapttkpghtapttkpapttkpkpkelap----- 804

Db 1097 rrsqssteaeptqdgagggsspsapatsvppspvptpaspsgpatlftsaf 1156
QY 1171 QLSPRQAGPPESLGVPAQAEALRGVPGGSITKGPSTKVPSDSAITVSGST 1230
Db 1157 gaalvgaarregwqnearrtflstfdagdedgdsqg-p9g--ppgprlrhsksid 1213
QY 1231 HGTADVLKGYITRIIGDPSRDLRGREDSLPKGHVIEYKKGHVLSYEGGSMVTQCS 1290
Db 1214 eg-----mfaeapylriesg-----gssggygayaag-----s 1241
QY 1291 KEDGRSSGPPHETAAPKRTYDMEGRGRAISSASIEGLM-----GRAI 1335
Db 1242 rayggsgsssaftsflppr--plvhlqkalpspqlalaaareralkesseggtq 1299
QY 1336 PPER-----HSPH--HLKQKHIR-----GSITQGI 1359
Db 1300 ppprpsprydappptllhshpsphshparhepvrlrhwgdparrtelgyraglsgqkal 1359
QY 1360 PRSYVEAQEDYLR-----EAKLLKREGTTPPPPSRDLTEAYKTQALGPLKPAHE 1412
Db 1360 tasppaarrrllhriptapgvvgplllqig-pepptphpgvskawrtaa----- 1407
QY 1413 GUVATVKEAGRSIHETPRELRHPE-LPLAPRLKEGSITQGTPLKYDTGASTTGSKKH 1471
Db 1408 -----pee-----perlphvrflen--cqarp-----ppagtrgsste 1439
QY 1472 DVRSLTGSPGRTFPVPHPLDVNMADARALERACYEESLKRSGCTASSSGSIARGAPVIVP 1531
Db 1440 d-----gpg--vppsprvl-----ptsptsrneenglllv- 1472
QY 1532 ELGKPRQSPLTYYEDHGAFAGHLPRGSPVTMPREPTRLOEGLSSSKASQDRKLJSTPRE 1591
Db 1473 ---lppapsvdvddg-----eflfaeplppllefsnsfek----- 1505
QY 1592 IAKSPHSTVPEHHHPHPISPYEHLLRGVSDVLYRSHIPLAFDPTSIPRGIPLDAAAAYYL 1651
Db 1506 ----pesplitpgpphplp-----dpps--patpipa----- 1531
QY 1652 PRHLAPNTPYHLYPYLIRGYPTAALENROTIINDYITSOOMHNTATAMAQRADMLR 1711
Db 1532 -----ppavaaappt--ldstassitsy----- 1553
QY 1712 GLSPRESSIALNYAGPRGIDLSQVPHLPVLVPTPGTPATAMDRLAYLPTAPOPFSSR 1771
Db 1554 -----dsevatltqgapaapg-----ppapgpapa---apabpapg----- 1589
QY 1772 HSSSLPSGGPHTLTKPTTSSSERORDRDREREKSIJLTSTTVEHAPIWRPCT 1831
Db 1590 -----gp---dppp9tdsgieevdsrssdhple-----tissastlssl-----sa 1628
QY 1832 EQSGSGSGSGSGSGSGSRPASHAHQSPISPRPTQDALQORPSVLNHTGMKIITAVE 1891
Db 1629 eggngtgvaggagvan-----gtellldtyayldgqafgsggtppg 1671
QY 1892 PSKPTVL-----RSTSTSSVPRPAATFPATHCPGLGTLGVYPTLMPEVLLPKEAPR 1944
Db 1672 pyppqlmtpsklirgalgtsgnlrpg-----psgglrdpvtpt-----spt 1712
QY 1945 VARPERPRADTGHFLAKPPARSGLEPASSPSKSGEPRLPPVPSGHATIARTPAKNLAP 2004
Db 1713 vs--vtgagtdgllalsacpbgpstagvaggp---vavepevpvpplpa--asslprkilp 1765
QY 2005 HHASDPDPAP--PASASDPHREKTS-QPFSIQELELSLGVHSGSSYSGEPVSPVSS 2061
Db 1766 weegppppppplpgplsqasalatvkslielskllqqfggss--taggalpwarggs 1824
QY 2062 PSLT--HDKG-----LPKHLELDKSHLEGELRPKQCPVKLGGEAAHPLRPLPESQFSS 2116
Db 1825 ggstdshggasyipertsslqrqlrsed-----sqtssl-----skpss 1864
QY 2117 S-----PLLOTAPGVKGHQRVWTLAQHISEVITQDYYRHHHPQQLSAPLPAPLYSF 2166

Db 1865 sifqnwkpplpplptgsgvsstaapatspsassasastr-hlqgvfemrppl---- 1920
QY 2167 PGASCPVLDLRRPPSDLYLPPDPDHGAPARGSPHSEGGKSRPEPNKTSV----- 2214
Db 1921 -----lrrapsllipasdh-----kvspaprpssllpilsppiypgl 1958
QY 2215 -----LGGEDGIEPV-SPPE-----GMTEPGHSRSVAVYPLLYRDGEOTEPSR 2256
Db 1959 fdirsstgtgaggstdpfapvfvpphpglsgglggalsgasrs-----lsptr 2006
QY 2257 MGSKSPGNT--SQPPAFFSK 2274
Db 2007 iislpdpkpf9akplgfwtk 2026

Search completed: September 8, 2001, 14:34:13
Job time: 17730 sec

Db 70 PSGKAQPHASVYVSEAGKDKGPPKSRVEELRTRGKTTTAAANFIDVTITRQIASDKDA 129
QY 2022 HREKTSKPFSTQELRLSLGYHSGSYSPGVEPVSPVSSPLTHDKGLPKHLELDKSH 2081
Db 130 RERGSSDSS-----SSLSSHRVETASDALEVISPASSAPPOEKQAOYQPDVMVKANQ 183
QY 2082 LEGELRPKQPGVKLGGEAAHPLHLR-----LHHRYSQOESPSPQOQPLPSSSEGGMQVPRTHRLIT 2134
Db 184 AENESTROYEGP-----LHHRYSQOESPSPQOQPLPSSSEGGMQVPRTHRLIT 234
QY 2135 LAQHSITVDYTRHHPOQLSAPLAPLYSPGA--SCPVLDLRRPPSDLYLPPD--- 2189
Db 235 LADHICQIITDFARN--QVPSQPTSTFTQSPALSSTPV---RTKTSRYSPEQSQT 289
QY 2190 --HGAPA-RGSPH-----SEGKRSPEPNKTSVLGGGEGIEPVSPPEGMTEPG-HSRSA 2240
Db 290 VUHPRGPRVSPENLVDRSGRSPKSPERSHI---PSEPIEPIPPQG---PAYHEKOD 343
QY 2241 VYPLLYRDEQTEPRMGSKSPGNTSQPAFFSKLTESNAMSVMKSKQKQINCKLTHNRN 2300
Db 344 SMLLSQRGVDPAEQRSRSPGSIYLPSPFTKL-ESTSPMVKSKKQEIFRKLNSGGG 402
QY 2301 EPEYNISQPGTEIFNMPAITGTLTYRSQAOVEHASTNMGLEAIRKALMGKYQWEE- 2359
Db 403 DSDMAAQPGEIFNLPAVTTSGAVSSRSHSPADPAS-NLGLIEDIRKALMGFDDKVED 461
QY 2360 -----SPPLSANAFNPLNASLPAAMPITTAADGRSDHLLTSPGGG-GKAKVSGRPSRK 2413
Db 462 HGVVMSHPV---GIMPGSASTSV-----VTSSEARDEGEPSPHAGVCKPKLINKSNRK 513
QY 2414 AKSPAPGLA--SGDRPPSVSVHSEDCNRRTPLTNRVWEDRPPSSAGSTPPFPNPLIMRL 2471
Db 514 SKSPIPGQSYLGTERTPSSVSVHSEGDYHROTP--GWAWEDRPSSTGSTQFPYNPLTIRM 571
QY 2472 QAGVNASPPPLPAGSGPL--AGPH---HANDEPKPLLCQYETLSDSE 2517
Db 572 -----LSSTPPTQIACAPSAITQAPHOQNRWIEREPAPLLSAQYETLSDSD 618

RESULT 2

PCT-US95-16311-5
; Sequence 5, Application PC/TUS9516311
; GENERAL INFORMATION:
; APPLICANT: Moore, David
; APPLICANT: Seoul, Wongi
; APPLICANT: Choi, Hwang-Sik
; TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street, Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16311
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/372,652
; FILING DATE: 13-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/246001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070

; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 619 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
PCT-US95-16311-5

Query Match 5.4%; Score 718; DB 5; Length 619;

Best Local Similarity 36.0%; Pred. No. 1,5e-33;
Matches 213; Conservative 76; Mismatches 212; Indels 90; Gaps 26;

QY 1975 PSKGSEPRLV-----PPVSGHATIAITPAKN--LAPHHASDPDPAPPASADP 2021
Db 70 PSGKAQPHASVYVSEAGKDKGPPKSRVEELRTRGKTTTAAANFIDVTITRQIASDKDA 129
QY 2022 HREKTSKPFSTQELRLSLGYHSGSYSPGVEPVSPVSSPLTHDKGLPKHLELDKSH 2081
Db 130 RERGSSDSS-----SSLSSHRVETASDALEVISPASSAPPOEKQAOYQPDVMVKANQ 183
QY 2082 LEGELRPKQPGVKLGGEAAHPLHLR-----LHHRYSQOESPSPQOQPLPSSSEGGMQVPRTHRLIT 2134
Db 184 AENESTROYEGP-----LHHRYSQOESPSPQOQPLPSSSEGGMQVPRTHRLIT 234
QY 2135 LAQHSITVDYTRHHPOQLSAPLAPLYSPGA--SCPVLDLRRPPSDLYLPPD--- 2189
Db 235 LADHICQIITDFARN--QVPSQPTSTFTQSPALSSTPV---RTKTSRYSPEQSQT 289
QY 2190 --HGAPA-RGSPH-----SEGKRSPEPNKTSVLGGGEGIEPVSPPEGMTEPG-HSRSA 2240
Db 290 VUHPRGPRVSPENLVDRSGRSPKSPERSHI---PSEPIEPIPPQG---PAYHEKOD 343
QY 2241 VYPLLYRDEQTEPRMGSKSPGNTSQPAFFSKLTESNAMSVMKSKQKQINCKLTHNRN 2300
Db 344 SMLLSQRGVDPAEQRSRSPGSIYLPSPFTKL-ESTSPMVKSKKQEIFRKLNSGGG 402
QY 2301 EPEYNISQPGTEIFNMPAITGTLTYRSQAOVEHASTNMGLEAIRKALMGKYQWEE- 2359
Db 403 DSDMAAQPGEIFNLPAVTTSGAVSSRSHSPADPAS-NLGLIEDIRKALMGFDDKVED 461
QY 2360 -----SPPLSANAFNPLNASLPAAMPITTAADGRSDHLLTSPGGG-GKAKVSGRPSRK 2413
Db 462 HGVVMSHPV---GIMPGSASTSV-----VTSSEARDEGEPSPHAGVCKPKLINKSNRK 513
QY 2414 AKSPAPGLA--SGDRPPSVSVHSEDCNRRTPLTNRVWEDRPPSSAGSTPPFPNPLIMRL 2471
Db 514 SKSPIPGQSYLGTERTPSSVSVHSEGDYHROTP--GWAWEDRPSSTGSTQFPYNPLTIRM 571
QY 2472 QAGVNASPPPLPAGSGPL--AGPH---HANDEPKPLLCQYETLSDSE 2517
Db 572 -----LSSTPPTQIACAPSAITQAPHOQNRWIEREPAPLLSAQYETLSDSD 618

RESULT 3

US-08-061-376-5
; Sequence 5, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Diabali, Malek
; APPLICANT: Sella, Lucia
; APPLICANT: Farry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
; TITLE OF INVENTION: TRANSLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles

STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/061.376
FILING DATE: 13-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9387
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)546-4737
TELEFAX: (619)546-9392
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3969 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-061-376-5

Query Match 2.94; Score 380; DB 4; Length 3969;

Best Local Similarity 18.09; Pred. No. 3.3e-13;

Matches 590; Conservative 357; Mismatches 1082; Indels 1250; Gaps 148;

Qy 51 SHLSPGSIQORRPSLLS-EFQGNERSQELHLRPSHSYLP-----ELGKSEMEFTES 105
Db 679 SPLHSGTFRDMHRSPLLRPRFT-----SEAHSRIFESVTLPSNRTSAGTSSSGVSNR 733
Qy 106 KRPLELLPDLRLPSPLLATGQAGSEDLTKDRSLTKGLEPVSPSPPHTDPELELYPP 165
Db 734 KKRKRVFSP---IRSEP---RSPSHSMRTSRGLSSLSSELSPLTPSSVSSLSISVSP- 785
Qy 166 RLSKEELLONDRVDREITVMEYQIISKKKQOQLEEAAPPEPEKPVSPPIESKURS 225
Db 786 -LATSALNPTFTFPHSHLT-----QSGSAEKQNRPKOTSAPAEFPSSSSPTPL----- 834
Qy 226 LVQIIYDENRKKAAAHRLLEGGLGQVELPLYNQPSDTRQYHENIKINQAMRKKLILYFK 285
Db 835 -----FPWFTPGSQTERGNKDKAPEELSKD----- 860
Qy 286 RRNHARKQWKQFCQRYDQLEALEKKVERIENPNRRRAKESKVREYEEKQFPEIRKQRE 345
Db 861 -ROADKSVEKDK-----SRERDREREKENRESREKR-----KKGSEIQSSSA 903
Qy 346 L-----QERMQRVQRGSGLSMSAAR-SEHVESEIIDG-----LS 380
Db 904 LYPVGRVSKERVVGEDVATSSSAKATGRKKSSSHSDGTDITSVTLGDTTAVTKILIKK 963
Qy 381 EOENLEKQMRQVLAVIPPMYLDADQORIKFINNGLMADPMKVIKDRQVMNMWSEOEKETF 440
Db 964 GRNLEKTNLDLGTAPSL-----EKEKTLCLTSPSSSTVK-HSTSSIGMSLAQADKLPM 1017
Qy 441 REKFMQHPKFNGLIASFLERKTVAEVCVLYIYTKKNENYKTVRRRSYRRRKSGSQOQQQQ 500
Db 1018 TDK-----RVASLLKKAQALC-----KIEKSKSL----- 1042
Qy 501 QOQOQOQOQOQWPRSSQEKEKEK-----EKEAEKEEKPEVE-- 539
Db 1043 -----KOTQPAQAGQSSSETSVRGPRIKHYCRRRAAVALGRKRAVFPDDMPTLSAL 1095
Qy 540 --NDKEDLLKEKTDGTDSCEDNDEKAVASKRGTANSQGRKKGRITRSMANFANSEAI- 596
Db 1096 PWEEKIL-----SSMGNDKSSIA--GSEDAEPLAPPIKIPVTRNKAPQEPVVK 1146

Qy 597 ---TPQSAELASMELINESSRWTEEMETAKKGLLEHGRNWSAIARMVSGSKTVSOCKNEY 653
Db 1147 KGRRSRCGQCPGQVPEDCGVCTNCLDKPKFG-----GRN-----IKQC----- 1187
Qy 654 FNYKRRQNLDEILQOHLKMEKERNARRKKKAPAAASEAAFPVVEDEMEASGVSN.713
Db 1188 CKMRKQNLQ--WMPSKAYLQAKAVKKKEKSKTS-----EKKDSKSSVVK 1235
Qy 714 EEMVEEAALHASGNEVPRGCSG-----PATVNNSSDTEIPSPH-----TEAAKD 761
Db 1236 ---VYDSQKPTPSAREDPAPKSSSEPPPRKPVVEKSEEGNVSAPGPEKSKATTPASRK 1292
Qy 762 TGQNGKPPATLGADGPPPPPT--PPRRTSRAPLETPASEATGAPTPPPAPPSPS-- 816
Db 1293 SSKQVSQPALVI-----PPQPTTGPRK--EVP-KTTPSEPKKKQPPPEPSPQSKOK 1344
Qy 817 --APPVVPKEKEEETAAAPVVEGE-----EOKPPA----- 847
Db 1345 KVAPRPSIPVKQPKPEKEKPPPVNKOENAGTNILSTLSNGNSSKOKIPADGVHRIKRVDF 1404
Qy 848 ---AEE-----LAVDTGKAE-----EPVKSCTEEAE 872
Db 1405 KEDCEAENWEMGGILITSVPIITPRVVCFCASSGHVEFVYVCVCEFFHFKLEE-NE 1463
Qy 873 GPAK-----GKDAEAAEATAE-----GALKAEKKE-- 897
Db 1464 RPLEDOLENWCCRCCKFCHVCGRQHOATKQLECNKCRNSYHPECLGPNYPTRKTKKKV 1523
Qy 898 -----GSGRATTAKSSGAPQDSSSA-----TCSADEVEAE 930
Db 1524 WICTKVRCKSCGSTTPGKWDQAWSHDFSLCHDCAKLFAGNFCPLCDKCYDDDDYDSEK 1583
Qy 931 -----GGKNRLLSRPSLLTPT-----GDPRANASPKQLDL 963
Db 1584 MMQCKDCDRWHSKCNLSDEMYEILSNLPESVATVNCNTERHFAEMLALEKELQISL 1643
Qy 964 KQ-----LKQRAAAIPIQVTKVHEPRDEAATKPAAPPAPPQNLQPS 1009
Db 1644 KQVLTALLNSRTTSHLLRYQAAPKPDLL-----NPETEESIPSRSSPEGPOPPVLTEVSK 1698
Qy 1010 DAPQPGSSPRGKSRPAPPADKEAFABAQKLDGDPQPCWTSGLFPVPVPRVVIKASPIA 1069
Db 1699 QDDQOP-----LDLEGVKKRMDQGNITSVLEFSDDDIVAKIIQAANS 1739
Qy 1070 PDPSAFSVAPGHPPLPLGLHDTAP-----VLP-----RPPTISNPPPLISS 1111
Db 1740 DG-----GOPEIKKANSVMKSFFFIOMERVEPWFVSVKSRFWEPNKVSNSGMLPN 1790
Qy 1112 AKHPSVL-----ERQIGA-ISQMSVQLHVPYSEHAKAPVGPVTMGLPLPMDPKKLA 1162
Db 1791 AVLPPSLDHNVAQWQEREENSHTEQPLMKKIIP-----APKPKGPEPSPPLHPPTPP 1846
Qy 1163 PFGVKQBSLSPRQAGPPESL-----GVPTAQEA----- 1192
Db 1847 ILSTDREDSRP--ELNPPPGIEDNRQCALCLTYGDDSDANDAGRLLYIGQNEWHVNCAL 1904
Qy 1193 -----SVLRGTALG-----SVPGGSITKGPSTRVPSDSAITVRSIT 1230
Db 1905 WSAEVEFDDGSLKNVHVAIRGKQLRCEFCOKPATV--GCCLISCTSN----- 1952
Qy 1231 HGTPADVLYKGTITR-----IIGEDSPSLDRGREDLSLPGHVI-----YEG 1272
Db 1953 -----YHFMCSRAKNCVFLDDKKVYQHRD--LIKGEVVPENGFEVFRVFDVEG 2002
Qy 1273 -----KGHVLSYEGMSVT-----QCSKEDGRSSSGPP 1301
Db 2003 ISLRKFLNGLEPENIHMIGSMITDCLGILNDLSDCEDKLPFGYQCS----- 2051
Qy 1302 HETAAPKPTYDMMEGRVGRATISSASIEGLMGRAIPPERHSPHHLKEQHHRISITOGIPR 1361
Db 2052 -----RVYSTTTPDARKCVYTKIVCEPRPVVPEDINSTVEHDENTIAHS-----PT 2099
Qy 1362 SYVEAQEDYLREAKLLKREGTPPPPPPSRDLTEAY-----KTOALGPLKLPKPAH 1411

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-255-62

Query Match 2.9%; Score 377.5; DB 1; Length 1064;
Best Local Similarity 21.7%; Pred. No. 8.7e-14;
Matches 353; Conservative 40; Mismatches 545; Indels 691; Gaps 84;

QY 694 AAFPVVDEMEASVSGNEEMVEAEALHASGNEVPRGECSPATVNNSSDTESTIPS 753
DB 23 AAHPFASD-PMGAHGPAGPK-----GAHGPAGPKG-APGAPPGSGRDPGPGP 70
QY 754 PHTAAKTGQNGKPPATLGADGP-----PGPTTPRRTSRAPTEPASEA---TG 804
DB 71 APGPAGPGSRGDPGPGAGPAGPGSGRDPGPGAP-----GAPG-PSGRGDPGPG 124
QY 805 APTP-----PPAPPSPAP-----PVVPEKEEETAAAPVVEEGEQK 845
DB 125 APGPAGPGSRGDPGPGAGPAGPGSGRDPGPGAP-----GAPGPGSRG-DGCP 176
QY 846 PAEELAVDTKAEPPVSECTEEAEEGA--KGDAEAAEATAGALKAEKKEGSGRA 903
DB 177 PGAPGAPPGSGRDP-----GPPGAPGAPPGSGRDPGPGAPGAPPGSGRDPGPP 231
QY 904 TTAKESSGAPQSDSSATCSADEVDEAGDKNRLLSRPSLLTPTGDPANASPKPLDL 963
DB 232 GAPGAPGP-----GSRGDPG-----PPGAPGAPPGSGRDPGPP--- 267
QY 964 KOLKORAAIPIQVTKVHEP-----PREDAAPTTP-APPAPPPQNLQPEDSA 1011
DB 268 -----GAHGPAGPKGAHGPAGPKGAHGPAGPKGAHGPAGPGSGRDPG 320
QY 1012 PQOPG-SSPRKSRSPAP-ADKEFAEABAKLPQDPCWTSGLFPV--PPREVIKASP 1067
DB 321 PGAPGAPPGSGRDPGPGAGPAGPGSGRDPGPP-----GAPGAPPGSGRDPGPP 375
QY 1068 HAPDP-----SAFSAVAPGHLPLG-----LHDTARPLVPRPTISNPPPLISSAKHPSVL 1118
DB 376 GAPGAPPGSGRDPGPGAGPAGPGSGRDPGPGAGP-----AGPPG----- 421
QY 1119 ERQIGAIQSGMSVQLHVPYSEHAKAPGV--TMGLPLPMDPKKAPFSGVKOEQLSPRG 1177
DB 422 -----SRG-----DPGPGAPGAPPGSGRDPGPPG-----AP-----GP 452
QY 1178 AGPPESLGPVTAQAEASVLRGALGVPGSGITKIPSTRVPSDSAITYRGSTHGTADV 1237
DB 453 AGPPSGRDP-----GPPGAPGAPPGSGR-----GDPGPP 483
QY 1238 LYKGTITRIIGEDSPRLDREDLSLPGHVIYEKKGHVLSEYEGMSVTQCKEDGRSS 1297
DB 484 -----GAPGAPPGSGRDPGPGAGPAGPAGP-----SRGD 515
QY 1298 SGPP--HETAAPKRTYDMMEGRVGRASISAEGLMGRAIPPERHSPHLKEQHHRGSI 1355
DB 516 PGPPGAHGPAGPK-----GAHGPAGPKGAHGPAGPKGAHGPAGP- 554
QY 1356 TQGIPTSVYEAQEDYLRREAKLLKREGTPPP--PPPSRDLTEAYTQALGPLKPAHEG 1413
DB 555 -----KGAPGAPPGSGR-----DGPPG 574
QY 1414 LVATVKEAGRSIETPREELRHTPELPLAPRLKEGSITQGTPLKYDTCASVTGSKKHVD 1473
DB 575 APGPAGPGS-----RGDPGPGAGP-----AGPPGSR----- 603
QY 1474 RSLIGSGRTFPVPHPLDMADARALERACYEESLSRPGTASSSGSGTARGAPVIVPEL 1533
DB 604 -----GDGP--PP-----GAPGAPPGP-----SRGDP----- 624
QY 1534 GKPROSPLTYEDHGAFFAGHLPRGSPVMTREPTPKLQGSISSSKASODRKLSTSPREIA 1593
DB 625 GPP-----GAPG-----PAGPPGSGRDPGPPGAPG-----PAGPP 654

QY 1594 KSPHSTVPEHHHPHPISPYEHLLRGVSGVDLYRSHIPFLAFDPTSIPIRGIPLDAAAYLPR 1653
DB 655 GSRGDPGPPGAPGAPG-----PGSRG-----DPG--PPG----- 682
QY 1654 HLAAPTYPHLYPYLIRGYPDTAALLENQTIINDYITISQOMHHNTATAMAOADMRLGL 1713
DB 683 --APGAPG---PG--SRGDPGPPG----- 700
QY 1714 SPRESSLANVAAGPRGIDLSQVPHLPVLVPP-----TGTATAMDRLAYLPTAPQPS 1769
DB 701 -----APGAPG-----PPGSRGDPGPPGA-----PGAPGPG 727
QY 1770 SRHSSSLSPGCGPHTLTKPTTTSSSERERDRDROREREKSIILTSTTVEHAPIWPR 1829
DB 728 SRGDPG--PGAP----- 744
QY 1830 GTEQSSGSSGSGGGGSSSR-----PASHSHA---HQHSPISPRQTQDALQORPSVLHN- 1880
DB 745 GSRGDPGPPGAPGAPGPGSGRDPGPGAGHGPAGPKGAHGPAGPKGAHGP- 803
QY 1881 TQMKGIITAVEPSKPTVLRSTSTSPVRPAATFPATHCPLGGLDGVYPTLMPEVLLPK 1940
DB 804 AGPKG---APGAPGPGSGR--DPGPPGAPGAPGPPGSGRDPG-----PP 843
QY 1941 EAPRVARPERPRADTGHAFKAPKPARSGLEPASSPKGSEPRPLVPPVSGHATARTPAK 2000
DB 844 GAPGAPGPPSGRDPG-----PPGAPG--PAGPGSGRDPG----- 878
QY 2001 NLAHHASPDPPAPPASASDPHREKTQSKPFSIQEELRSLGYHGSSYSGEVPVSPVS 2060
DB 879 -----PGAPGAPGPGSGRDP-----GPPGAP--GPAG 904
QY 2061 SPSLTHDKGLPKHLELDOKSHLEGLRKPQOPVKLGGAHLPHLRPLPESQPS--SP 2118
DB 905 PPGSRGDPGPPG-----APGAPGPPGSGRDPG--PPGAPGAPGPPGSGRDP 948
QY 2119 LIQTAPGVKGHQRVVTLAQHISEVITQDTRHHPOLLSAPLPA-----PLYSPPG 2168
DB 949 GPPGAPGAG-----PPGSRGDPGPPGAPGAPGPPGSGRDPG-----PG 988
QY 2169 ASCPVLDRRPPSDLYLPPPDHGA- RGSHPSEGGKRSPEPNKTSVLG-GGEDGIEPVS 2226
DB 989 APGPA---GPPSGRDPGPP--GAPGAPGPPGSGR-----DPGPPGAHGPAGPKGAHGA 1038
QY 2227 PPEGMTPEG 2235
DB 1039 GPKGAMDPG 1047

RESULT 5
US-09-041-886-23
; Sequence 23, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharro
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polyepitopes and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

;; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
;; FILE OF INVENTION: PREPARATION THEREOF
;; CURRENT APPLICATION NUMBER: US/09/219,849
;; CURRENT FILING DATE: 1998-12-23
;; NUMBER OF SEQ ID NOS: 50
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 5
;; LENGTH: 960
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
;; OTHER INFORMATION: amino acid sequence
US-09-219-849-5

Query Match 2.8%; Score 365; DB 4; Length 960;
Best Local Similarity 21.3%; Pred. No. 4e-13;
Matches 337; Conservative 33; Mismatches 490; Indels 720; Gaps 78;
QY 726 ASGNEVPGECSGATVNNSSD--TESIPSPHTEAAKDTGQNGPKPPATLGDGPP----- 779
DB 11 AHGPAGPKG-APGPAGPGSGRDPGPPGAPG-----AGPPGSRDPGPPGAPGPPGSRD 65
QY 780 PGPP-----TPPRTSRAPTEPTPASE-----ATGAPT--PPA---PPSPSAPP 820
DB 66 PGPPGAPGPPGSRDPGPPGAPGPPGSRDPGPPGAPGPPGSRDPGPPGAPGPPGAPG 125
QY 821 VVPKEEKEETAAPVVEEGEQKPAEELAVDTGKAEEPVKSECTEEAEGPAKGD 880
DB 126 AGPPGSRDPGPPGAP-----GPAGPPGSRDPGPPGAP-----GPAGPPGSRDP 168
QY 881 EAAETAEGALKAERKEGSGRATTAKSSGAPQSDSSATCSADEVDEAEGDKNRLSP 940
DB 169 GPPGAPGPPGSRDPGPPGAPGPPGSRDPGPPGAPGPPGSRDPGPPGSRDP----- 221
QY 941 RPSLTPTGDPANASPOKLDLKKLORAAAIPTQVTKVHEPPREDAAP-----T 992
DB 222 -PGAHGAPG- PKAHGAPG-----KGAGHAGPKGAGHAGPKGAGPAGPPGSR 270
QY 993 KPAPPAPPPQNLQESDAPQPGS-SPRGK--SRSPAPPADKEAFAEAKLPDGPQW 1049
DB 271 DPGPPGAPGAPG-PPGSRDPGPPGAPGPPGSRDPGPPGAPGPPGSRDPGPP----- 326
QY 1050 TSGLEPPVPPREVIRKASHPADPSAFYAPGHPPLPLGLHDFARVLPPTISNPPPLI 1109
DB 327 --GAPGAPG-----PGSRDP-----GPPGAPGPPGSRDPGPPGAPGAPGPP----- 369
QY 1110 SSAKHPVLERIGAIQSGMSVOLHPVYSEHAKAPVPTMGLPLPMDPKLAPFSGVKQ 1169
DB 370 -----GSRDPG-----PPGAPGAPG-----PGSRDP----- 391
QY 1170 EQLSRGAGPPESLGVTQAQASVLRGALSGVPGSGITKIPSTRVPSDAITYRGI 1229
DB 392 -----GPPGAPGAPGPPGSRDP-----GPPGAPGAPGPPGSRDPGPP----- 428
QY 1230 THGTADVLTKTITRIGEDSPSLDRGREDLSLPGHVIYEGKGHVLSEGGMSVTQC 1289
DB 429 --GAP-----GPAGPP-----GSRDPGPPG-----APGAPG----- 452
QY 1290 SKEDGRSSSGPP--HETAAPKRTYDMMEGRVRAISSASIEGLMGRAPIPERHSPHLLKE 1347
DB 453 --PPSRDPGPPGAPGAPGPK-----CAHGPAGPKGAGHAGPAGPKG 490
QY 1348 QHHIRGSITQIGIPRSYVEAQEDYLRREAKLLKREGTPPP--PPPSRDTEAYKTOALGPL 1405
DB 491 AHGPAG-----KGAPGAPGPPGSRD----- 511
QY 1406 KUKPAHEGLVATVKEAGRSIHPIPRELHRHTPELPLAPRLKEGSITQCTPLKYDTGAST 1465
DB 512 -----PGPPGAPG-----AGPPGSRDPGPP- 532

QY 1466 TGSKKHVRSLIGSPGRTFPPVHPLDVMADARALERACYEESLKSRPGTASSSGSIARG 1525
DB 533 -----GAPGAPGPP-----GSRDPG 547
QY 1526 APVIVPELKGKQPSPLTYEDHGAPFAGHLPGRGSPVYTMREPTRLQEGSLSSKASODRKL 1585
DB 548 PPGAPGAPGPPG-----RDPGPPGAPG-PPGPPGAPG-PPGPPGAPG-PPGPPGSRD--- 596
QY 1586 TSTPREIAKSPHSTVPEHHHPHPISEYHLLRGVSGVDLRYSHIPLAFDPTSPRGIPLDA 1645
DB 597 -----PGPPGAPGAPG-----PGSRDPG-PPG----- 618
QY 1646 AAAYLPRHLAPNPTYPHLYPPYLINGYDPTAALENROTIIINDYITTSQQMHHTATAMAQ 1705
DB 619 -----APGP----- 622
QY 1706 RADMLRGLSPRESSALNAYACPRGIIDLSQVPHLPVLVPTPGTPTATA---MDRLAYLP 1762
DB 623 -----AGPPGSRD-----PGPPGAPGAPGPPGSRDPGPP 651
QY 1763 TAPQPFSSRHSSPLSPGPGPHTLTKPTTTSSSERDRDRDREREKSIILTSTTVE 1822
DB 652 GAPGAPGPPGSRDPGPPGAP----- 671
QY 1823 HAPIWRPCTEQSSGSSGSGGGSSSR---PASHSHA---HOHSPISPTQDALQORP 1875
DB 672 -GPAGPPG- RDPGPPGAPGAPGPPGSRDPGPPGAGHAGPAGPK----- 720
QY 1876 SVLHNTGMKGIITAVEPSKPTVLRSTSSVVRPAATFPFATHCPGLGGLDGVYPTLMEP 1935
DB 721 -----GAGH-----PAG--PKGAGPAG----- 736
QY 1936 VLLPEAPRVARPERPRADTGHAFKAPKPARSGLEPASSPSKGE-----PRPLVPPV 1988
DB 737 ---PKAGCPAGPPGSR-DPG-----PPGAPG--PAGPP--GSRDPGPPGAPGPPGPPG 782
QY 1989 SGHATARTPAKNLAPHHASP-----DP-----PAPPASADPHREKTSKPFISQIELE 2037
DB 783 S-----RDPGPPGAPGAPGPPGSRDPGPPGAPGAPGPPGSRDP----- 820
QY 2038 LRSGLYHGSYSYSPGVE-PVSPVSPSLTHDKGLPKHLELDKSHLEGLRKPQGPVKL 2096
DB 821 -----GPPGAPGAPGPPG-----RDPG-----PPGAPGAPG 848
QY 2097 GGEAAHLPLRPESOPSS-SPLLOTAPGVKGQRVVTLAQHISEVITQDTRHHHPQOL 2155
DB 849 PGSRDPGPPGAPGAPGPPGSRDPGPPGAPGAPGAPG-----PPGS 885
QY 2156 SAPLPAPLYSFFGASCPVLDLRRPDSLYLPPPDHGAPARGSPHSEGGKRSPEPNKTSVL 2215
DB 886 RDPGP-----PGAPGPA---GPPGSRDPGPP--CAPGAPG---PGSRDPGPP----- 924
QY 2216 GGGEDGIEVPSPPEGMTEPG 2235
DB 925 -PGAPG--PAGPP-GSRDPG 940

RESULT 7

US-07-853-913-4
; Sequence 4, Application US/07853913
; Patent No. 533839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; TITLE OF INVENTION: Neuroepithelial Tumors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.

RESULT 8

US-08-963-825-18
; Sequence 18, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
US-08-963-825-18

Query Match 2.7%; Score 351; DB 3; Length 1341;
Best Local Similarity 20.1%; Pred. No. 3.8e-12;
Matches 350; Conservative 106; Mismatches 573; Indels 714; Gaps 81;
QY 722 EALHAGNEVREGCGATVNNSSDTSPHTEAAK-DTGQNGKPPA-TLGADGPP 779
DB 77 ETKNCPGAEVGECC-PVCPDGSSTDDQETTVGEGRGDTGPRGPGAGPPGRDGIP 135
QY 780 -----PGPPTPRRTSRAPTEPTASEATGATPP-----PA 811
DB 136 GQGLPGPPGP-----GPPGPGLGNGFAPQLGYDEKSTGIGSYPG 178
QY 812 PPSAP-----PPVVKKEEETAAAPPVEE-----EOKPPAAEELAVDTGKA 858
DB 179 PMGPSRGLPGPPGAGPZGFZ-----GPPGZPGZPGASGPMGPRGPPGKBGBBGA 234
QY 859 EEPVKSETEAECPAKDAEAAEAAGALKAEKEGGSGRATTAKSSCAPODSS 918
DB 235 GKPR-----PGZRPFGPGZGANGLPFTA--GLPMKGHGFGSLBGAKBAGPAG---- 283
QY 919 ATCSADEVEAEGGDKNLLSPRPSLLTPTGDPANASFOKPKPLDLKQLKQRAAAIPPIQV 978

DB 284 -----PKGZPGS----- 290
QY 979 TKVHEPPREDAAPTKPAAPPQNLQESDAPQPGSGSPRCKSRSPAPPADKEAFAAE 1038
DB 291 -----PGZBAGPGZMGP-----GPKNSGEPGAGSGK----- 318
QY 1039 AOKLPDPPCWTSGLPFPVPPREVIVKASPHADPDPSAFSAPPCHPLPLGLHDTARVLP 1098
DB 319 -----GD--TGAKGPPGVQ-----GPPGAGEGKRGARCEPCTG-----LPG 358
QY 1099 PPTISNPPPLISSAKHPSVLERQIGAISQMSVOLHVPIYSEHAKAPVGPV--TNGLPLPMD 1157
DB 359 PPG-----ERG-GPGSRGF-----PGADGVAGPKGAGERGSGPAG 394
QY 1158 PKKLAPFSGVKOEQLSPRGQAGPESLGVPTAQEASVLRGALSGVPGSITKIGISTRV 1217
DB 395 PKG-----SP-GEAGRPEAGLPAGKLTGSPGS-----PGPDGKTGPPG--- 433
QY 1218 PSDSAITYRGSIHTGTPADVLYKGTITRIIGEDSPSRLDRGREDLSL-----PKGHVITYEGK 1273
DB 434 PAGQ-----DGRP-----GPPGPPGA-RGOAGVMGFPKGAAGEPCK 470
QY 1274 KGHVLSYEGM-----SVTQCKEDGRSSGPPHETAAPKRTYDMMGRVGRASSASIE 1328
DB 471 AG-----ERGVPGPAGVGPAGKDGEGAGAGPP-GPAGPAGE-----RGQGGPA-GSPGFQ 519
QY 1329 GLMGRAPP-ERHSPHLLKEQHHRIGSITQIGIPRSVVEAQEDYLREALLKREGTPPPP 1387
DB 520 GLPGPAGPGEAGKPE-----OGVPDGLGAPGSPGARGERFPGERGVQGP 567
QY 1388 PPSRDLTEAYKTOALQPLK--PAHEGLVATVKEAGRSIHEIPRELRHTPELPLAPR 1445
DB 568 GPA-----GPRGANGPNDGAKG---DAG-----APCAP-- 594
QY 1446 LKEGSITQGP-----LKYDTGASTTGSKKHD-----VRSLSGSPGR-----TFP--PVH 1488
DB 595 ---GS--QGAPGLQMPGERGAAGLPKPKDRGDAGPKGADGSPGDRGLTGTPGPG 649
QY 1489 PLDMADARALERACEESLSKRPCTASSSGSARGAPVIVPELCKPQSPITYEDHGA 1548
DB 650 PAGAFD-----KGESGSPGAGTARGAFGDRGEPGP--GPAGFA--GP 692
QY 1549 PFAGHLPRGSPVTMREPTRLQEGSLSSSKASQDRKLTSTPREIAKSPHSTVPEHHPHI 1608
DB 693 PGAD-----GQCAKGEFGDAGAKD-----ACPPGAPGAPGPGPI 729
QY 1609 -----SPYEHLRLRGVSDLYRSHIPLAFDPTSIPIRGIPLDAAAAYLPHRLAPNPTYPHL 1664
DB 730 GNVGAPGAKGARGAGP-----PGATGFPGAAGRVP---PGPS-GNA 768
QY 1665 YPPYLIRGYPDTAALNROTIINDYITSOQMHNTATAMAQADML-----RGL 1713
DB 769 GPP-----GPPGAGKEGKGPRGE-----TGPAGRPEVGGPPGPPAGEKGS 812
QY 1714 SPRESSLALNAAAGPRGIIDLQVPHLPVLVPPT--PGPATAMDRALVLTAPQPFSSR 1771
DB 813 PGADGPAGAPGTGPGQIAGQGVGLPCQGRGERFGLPGPSG-----PKQGPSGAS 867
QY 1772 HSSSPLSPGGTHLTKPTTTSSEERDRDRDREREKSIILTSTTTVEHAPIWRPOT 1831
DB 868 GERGPPGMPGGLAGPGGESGRE-----GAPGA 896
QY 1832 EGSSSSSSSSSSGGG-----GSSSRPASHAHQHSPISPRTQDALQORPSVLHNTGMKI 1886
DB 897 EGSPGRDGSAGKGRGETGAPGFCAGXAGAPGVGP-----AGKSD 941
QY 1887 ITAVPESKPTVLRSTSTSSVPRPAATFPATHCPLGGTLDGVTPLMEPVLLPKEAPVA 1946
DB 942 RGETPAGP-----AGPVGPAGARGP-----A 963
QY 1947 RPERPRADTGH-----AFLAKPPAKSGLEPASSPSKSGSRPLVPPVSGH 1991

Db 964 GPQGRGDKGETGEQGRGKIKHGRGFGSLQGGPPGPGSGEQPGSAGSP----- 1013
Qy 1992 ATIAITPAKNLAPHASDPDPAPPASADPHREKTQSKPFSIQEELSLGVHGSSSPE 2051
Db 1014 -----AGPRGPGSAGAPGKDGUNGLPPI-----GPPGRGRTG 1048
Qy 2052 GVEPVSPVSSPLTHDKGLPKHLELDKSHLEGRPKQPGVKLGGAHLPLRLPE 2111
Db 1049 DAGVPGPCPG-----PPGPPGPPSAGFDFSLP----- 1078
Qy 2112 SOPSSPLQTAGVKGHQ-----RVVTLAHSIVITQDYTR 2149
Db 1079 -Op-----POEKAHKGGRYRADDANVVRDRDLEVDTLKLSQOQIENIRSPGKR 1128
Qy 2150 HPOOLSAPLAPLYSFFCASCPLDLRRPPSDLYLPPDHCAPARGSPHSGGKRSEPE 2209
Db 1129 KNPAR-----TC--RDLKWCSDW-----KSGEYWDIP 1154
Qy 2210 NKTSLVGGGEGIEVPSPPEGMTEPGHRSAYVPLLYRDGEQTEPSRMG-----SKSPGN 2264
Db 1155 NQ-----GCNLDKAVFCNME-----TGTCVYP-----TQPSVAKKNWYISKNPKD 1196
Qy 2265 TSQPAFPSKLTES-----NSAMVSKKOBINKLTHNNEPEYNIS 2307
Db 1197 KDRHYMGESMTDGFQFEYGGGSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQ 1256
Qy 2308 QPG 2310
Db 1257 QTG 1259

RESULT 9

US-07-741-940-7
Sequence 7, Application US/07741940
Patent No. 5352775

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERTS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07741,940
FILING DATE: 19920109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-07-741-940-7

Query Match 2.7%; Score 350.5; DB 1; Length 2842;
Best Local Similarity 17.8%; Pred No. 1.1e-11;
Matches 473; Conservative 357; Mismatches 1007; Indels 821; Gaps 109;

Qy 50 ASHLSPGSIQIP--QRRRPSLLSEFQ-----PGNERSQELHLRPSHSY 91
Db 739 ANINSPGSSLPFLHVRKQKALEAELDAQHLSFTDNIDNLSPKASHRSKQRHKQSLYGDY 798
Qy 92 LPELGKSEMEFIESKRPRLELLPOPLLPSPLLATGQAGSEDLT---KDRSL----- 141
Db 799 VFDTRNRHDDNRSDNFTGNMTVLSPLYLNTTVLPSSSSRGSLDSSRSKDEKSLERERGI 858
Qy 142 TGKLEPVPPSPPHDPELELVPPRLSKEELIQNMDRVDREITMVEQOQISLKKKQOOLE 201
Db 859 LGNYHPATEN-----PGTSSKRGL-----QISTTAAQIAKV-----ME 891
Qy 202 EEAAPPEPEKVPSPPIESKHSRLVQIYYDENR--KKAEEAHRILEGLGPOVELPLYNQ 259
Db 892 EVSAIHTSQEDRSSGSTTE-----LHCVTDERNALRRSSAAH-----THSNTYN- 935
Qy 260 PSDTRQYHENIKINQAMRKLLILYFKRRNHARKQWKQFCQRYDOLMEALEKKVERIENN 319
Db 936 --FTKSENSNRTCSMPYAK---LEYKRSSN-----DSL-----NSVSSSDGY 972
Qy 320 PRRRAKESKVRREYKEK-----QFP-EIRKQRELOERMOSRVGQSGSLMSAAARSE 369
Db 973 GKRGOMKPSIESYSEDESCKFCSYQYPADLAHKTHSANHMDNDGELDTPINYSLKYSD 1032
Qy 370 HEVSEIIDGLSEQENLEKOMQLAVIPMLYDADQORIKFINMGLMADPMKVYKDRQVM 429
Db 1033 -----EQLNSGRQ-----SPSONERWARPKHIIEDIEIKSQSQRSR 1068
Qy 430 NMWSPQE--KETFRKFMQHPKNEFLIASFLERKTVACVLVLYLTKKNNYKSLVRSY 487
Db 1069 NOSTTYPVYTESTDDKHLKFPHPFG-----QQECVSPYRSRGANGSETNVRGNSH 1118
Qy 488 RRRGKSQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 541
Db 1119 ---GINQNVQSQCQEDDYEDDKPTNYSERYSEEQHEEERPTNYSIKYNEEKRHVDP 1175
Qy 542 KEDLLKEKT-----DTSGEDN-DEKPAVASKGKGTANSQQRGRTRISMA 587
Db 1176 IDYSLKYATIDIPSSOKQSFSSKSSGQSKTEHMSSENTSTPSSNAKRONQLHPSSA 1235
Qy 588 NEANSEEAITPQQAELASLAMELNESRWTEEMETAKGLLEHGRNWSAIRMVSKTSVS 647
Db 1236 QSRSGQ-----PQKAATCKVSSINQETIQTYCVEDTP-----ICFSRCSLSLS 1280
Qy 648 QCKNPFYNYKKRQNLND--EILQOHKLKMEK--ERNARKKKKAPAAASEAAFPVWVEDE 703
Db 1281 SAEDEIGCNQITQEADSANTLQIAEIK-EKIGTRSAEDPVSEVPVAVSOH-----PRTSS 1334
Qy 704 EMEASGVSGNEE-----EMVEEAALHASGNEVPGE-----CSGPATVNNSSDSTEPS 753
Db 1335 RLQSSLSSESARHKAVEFSSGAKSPSKGAQTPKSPPEHYVQETPLMFSTRCTSVSSLDS 1394
Qy 754 PHTAAKDTGQNGPKPPATLGADGP-----PGPPTPPRTSRAPIEPTPASEATGATP 808
Db 1395 FESRSIASSVQSEPCSGMWGIIISPSDLDPSPGOTMPPSRSK-----TP 1438

QY 809 PPAPSPSPVPPVVKKEEETAAAPPVEEGEQKPPAAEELAVDTGKA---EPPVKSE 865
Db 1439 PP-PPQTAQTKREVPKNK-----APTAEKRESGFKQAANVAQVQVLPDADLLHF 1490
QY 866 CTEEAEEG-----PAKGDAAEAATAGALKAEKKEGGGRATTA---KSSGAPQDS 915
Db 1491 ATESTPDGFCSSLSALSDEFFQOKVELRIMPVQVQNDNGNETESQPKESNENQK 1550
QY 916 DSSATCSA--DEVDEAEGD---KNRLSPRPRLTPTGDPANASPOKPLDLKOLKOR 969
Db 1551 EAEKTIIDSEKLLDDSDDDIIELECIISAMPT-----KSSRAKAP-----AOT 1596
QY 970 AAAIPPIQVTKVHEPPREDAAPTKAPPAPPONILQSPEDAPQPGSS--PRGKSRSPAP 1028
Db 1597 ASKLPPPVARKPSQLPVYKLLPSQ-----NRLQPKHVSFTPGDDMPR----- 1639
QY 1029 PADKEAFAEAQKLGCDPPCWTSGLPFPVPPREVYKASPHADPSAFSTAPPCHPLPLGL 1088
Db 1640 -----VYCVGEG-----TPINFSTA-----TSL 1656
QY 1089 HDTARVLPRPPTISNPPLISSAKHPSVLERQIGAISQMSVOLHVYPYSEHAKAPVGPV 1148
Db 1657 SDL-----TIESPP-----NELAAGEVGRGAQGEFEKRDITFT 1691
QY 1149 TMGLPLPMDPKKLAPFSGVYKQQLSPRGQAGPPESIGVP-----TAQEAASVLRGTALGSV 1203
Db 1692 -----EGRSTDEAQQGKTSSTVTPELDDNKAEEGDILAECINSAM 1731
QY 1204 PGK-----SITKGPITSRVPDSAITVR-----GSITHGTPADVLTKGTITRI 1246
Db 1732 PKGSHKPPFRVKKIMDVQOQASASAPKNQLDGKKKPTSPVKPIQONTYRTVRVK- 1790
QY 1247 IGEDSPRLDRGREDSPKGVHVIYEGKGVHLSYEGGMSVTOCKED---GRSSSGPPE 1303
Db 1791 -NADSKNNLNAERVS DNK-----DSKONLNNKSKDFNDKLPNNEDRVGRGFAFDSPHH 1844
QY 1304 TAAPKRTYDMEGRVGRATISSAIEGL-----MGRAIPPERHSPHLKQHHIRGSI 1355
Db 1845 -----YPIEGTPYCFNRNDSLSLDFDDDDVLSR-----EKAELRAKENKESEAKV 1893
QY 1356 TOGIPRSVYEAQEDYLREAKLLKREGPPP-----PPPSROL-----TEAKTO 1400
Db 1894 TSHTELTSQQQSANKTQAIKQPINRGQPKPILOKQSTFPQSSKDIPIRGAATDEKLQNF 1953
QY 1401 ALGPKLKAHGLVATVKEAGRSITHEPRELRHT-----PELPLAPRLKEGSI 1452
Db 1954 AIENTPVCFSHNSLSLSIDIOENNNKENEPKETEPPDSQGEFSKPOAGYAPKSFIV 2013
QY 1453 QGTPLKYDTGASTG-----SKHDRVRLSIGSPGR 1482
Db 2014 EDTVPCFSNRNSSLSSIDSDEDDLQECITSSAMPKKKPSRLKGDNEKHSRNMGGILGE 2073
QY 1483 TFPVPHLDVMADARALERACYEESLSKSPGTASSGGSIARGAPVIVPELCKPQSPIT 1542
Db 2074 -----DLTLDLKDIOQDPDEHGLS---PDSENFWDKAIQEGANSIVSSL----- 2114
QY 1543 YEDHGAPFAGHLPR-----GSPVTREPTPRLOEGSLSSKASQDRK 1584
Db 2115 ---HQAACAAACLSRQASDSDSLSLSKISIGSGFHL---TPDQEEKFTSNKG----- 2163
QY 1585 LUSTPREIAKSPHSTVPEHHPHIPISPEYHLLRGV--SGVDLYRSHIPLAFDPTSIPIRGIP 1643
Db 2164 ----PRILKPGKSTL-----ETKKIESESGIKGGKVKYS----- 2196
QY 1644 DAAAYYLPRHLAPNTYPHLYPPVLYRCYPTDALENQTIINDYITSQOMHHNTATAM 1703
Db 2197 -----LITG-----KVRNSNEISQOM----- 2212
QY 1704 AQRADMLRGLSPRESSLALNYAAGRPIDLSQVPHLPVLVPPPTGCTPATAMDRLAYLPT 1763
Db 2213 -----KQPLQANMPSISRG-----RTMTHIFGV----- 2235
QY 1764 APQPFSSRHSSPLSPGPTHLTKPTTSSSRERDRDRDRDREREKSIILSTTT--- 1820

Db 2236 ----RNSSTSPVSKKGGPP-LKTPASKSPSEQ-----TATTSPRG 2272
QY 1821 -----VEHAPITWRPCTEASSSGSSGGSSGSSPASHSHAHQHSPISPRTDALQOR 1874
Db 2273 AKPSVKAELSPLVARQ--TSQIGSSKAPSRGSRDSTPSRPAQOPLSRPIQSPGRNSI--- 2328
QY 1875 PSVLHNTGKGIITAVEPSKPTVLRSTSTSSPVVRPAATPPATHCPLGGTLDGVVP--- 1930
Db 2329 -----SPGRNGI---SPPNKLSQLPTTSPSTA-----STKSSGSKMSTSPGRQM 2372
QY 1931 ---TLMPEVLLKPEAPVPRPERPRADTGHAFILAKPARSGLEP-----ASSPSKSGE 1980
Db 2373 SQONLTQKGLSKNASSI-----PRSESASKGLNOMNGNANKKVELSRMSTSKSGSE 2427
QY 1981 ----PPLVPPVSGHATIAITPA-KNLAPHHASPDPPAPASADPHREKTSQKPSIQE 2035
Db 2428 SDRSERPVLRQSTFTKEAPSTLRKLEBASAFESLSSPSRSPASPTRSQAOPT-PVLSPS 2486
QY 2036 LELRSLGYHGSSVSPGVPVSPVSPSLTHDKGLPKHLEELDKSHLEGELRPKQPCPVK 2095
Db 2487 LPMDSLTH-SSVQAGGWKLPNLSPTTEYNDGRPAKRRHDIARSHSESPSL- ---PIN 2541
QY 2096 LGG--EAAHLPHLRPLPESQSSPLLQTPAGVKGHQRVVTLAQHISEVITODYTRHPQ 2153
Db 2542 RSGTWKREHSKH-----SSSLPRVSTWRTGSSSSILSASSESESEKAKSEDEKH--- 2590
QY 2154 QLSAPLAPLYSPGASCPVLDLRRPPSDLYLPPPHGAPARGSPHSEGGKRPENKTS 2213
Db 2591 -----VNSISGTK-----QSKENQVSAKGTWKIKENEFSPNTNSTS 2626
QY 2214 --VLGGEGDIE-----PVSPEGMTEPGHSRSAYVPLLYRDGEOTEPSRMGSKSPGNT 2265
Db 2627 QIVSSGATNGAASKTLLIYOMAPAVSKTEDVWVRIEDCPI-----NNPRSGRSTGNT 2678
QY 2266 SOPPAFFSKLTESNAWVSKKQKOEINKKLNTNRNEPEYNISQPGTEIFNMPAITGTGLM 2325
Db 2679 --PP-----VIDSVSEKANPNIKDKDNQAKQNV-----NGSV 2710
QY 2326 TYRSQAVQEHASTNMGLEALIRKALMGKVDQWEEESPPLSANAFNPLNASASLPAAPIA 2385
Db 2711 PMRTVLENLRLSFIQVDAPDQKGTEIKPCQNNPVVSETN-----ESSIVERTPFSS 2763
QY 2386 ADGRSDHTLTSPGGGKAKVSG---RPSRKAQSPAGLASGDRPPSVSVHSEGDGNCNR 2442
Db 2764 SSS-SKH--SSPSGTVAARVTFENYNPSPRKSAD-----STSRPSQIP----- 2805
QY 2443 TPLTNRVWEDRPSSAGST 2460
Db 2806 TPVNNT-KKRSKTDST 2822

RESULT 10

US-08-289-548A-7

; Sequence 7, Application US/08289548A

; Patent No. 5648212

; GENERAL INFORMATION:

; APPLICANT: ALBERTSEN, HANS

; APPLICANT: ANAND, RAKESH

; APPLICANT: CARLSON, MARY

; APPLICANT: GRODEN, JOANNA

; APPLICANT: HEDGE, PHILIP J.

; APPLICANT: JOSLYN, GEOFF

; APPLICANT: KINZLER, KENNETH

; APPLICANT: MARKHAM, ALEXANDER F.

; APPLICANT: NAKAMURA, YUSUKE

; APPLICANT: THLIVERIS, ANDREW

; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

; NUMBER OF SEQUENCES: 102

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Banner & Allegretti, LTD

; STREET: 1001 G Street, NW

Qy	1483	TFPPVPHLVMDARALEPRACYEESLSKBPCTASSSGGSIARGAPVIVYBELGRPQSP	1542
Db	2074	-----DLTKDIORPDSEHGLS--PDSENFDMKAIQEGANSVSL-----	2114
Qy	1543	YEOHGAFFAGHLPR-----GSPVTWREPTPLQGSILSSSKASQDRK	1584
Db	2115	---HQAANAACLSRQASSDSDLSLKSGISIGSPFHL---TPDOEKPTSNKG---	2163
Qy	1585	LTSTPREIAKSHUSTVPEHHHPHPISPYEHLLRGV--SGVDLYRSHIPLAFDPTSI	1643
Db	2164	---PRILKPGEKSTL-----ETKKIESEKINGKGGKVKYS-----	2196
Qy	1644	DAAAAYVPHLAPNPTYPHLYPPVPLRGYPDTAALENROFIINDYITLSQMHNTATAM	1703
Db	2197	-----LITG-----KYRNSETISGOM-----	2212
Qy	1704	AQRMDLRGLSPRESSALNYAAGPRGIDLSQVPHLVPVPPGTPATAMDRLAYLPT	1763
Db	2213	-----KOPLOANMPSISRG-----RTMIHIPGV-----	2235
Qy	1764	APQFSSRHSSPLSPGGTHLTKTPTTSSSERERDRDRDRDREREKESILTSITT---	1820
Db	2236	---RNSSSTSPVKKKPP-LKTPASKSPSEGO-----TATTSPRG	2272
Qy	1821	---VENAPIWRPTEQSSGSSGGGGSSRRPASHAHQHSPISPRQDQLQOR	1874
Db	2273	AKPSVSELSPVARQ--TSQIGSSKAPSRSGSRDSTSPRPAQPLSRPIOSPGENSI---	2328
Qy	1875	PSVLNHTGKGIITAVEPSKPTVLSTSTSPVRPAATFPFPAATHCPILGGTLDGYVP---	1930
Db	2329	---SPGRNGI---SPPNKLSQLRSTSPSTA-----STKSSGSGKMSYTSPCRQM	2372
Qy	1931	---TLMPEVLLKPEARVARPERPRADTGHAFKAPPARSGLEP-----ASSPKSGSE	1980
Db	2373	SQONLTKQGLSKNASSI-----PRSESASGLQNMNGNANKKVELSRMSSTKSGSE	2427
Qy	1981	---PRPLVPPVSGHATIARTPA-KNLAPHHASDPDPAPPASADPHREKTQSKPFIOE	2035
Db	2428	SDRSERPVLVROSTFTKEAPSPPLTKRKLEESASFESLSPSRPASPTRSQAQT--PVLSPS	2486
Qy	2036	LELRSLGYHGSSVSPGVEPVSPVSSPSLTHDKGLPKHLELDKSHLEGBLRKQPCPVK	2095
Db	2487	LPDMSLSTH--SSVQAGGWKRLPNLSPITETNDGRPAKRHDIAHSHSESPSRL-----PIN	2541
Qy	2096	LGG--EAAHLPHLRPLPESQSPSSPLQATAPGVKGHVVTFLAQIHSEVITQDYTRHHQP	2153
Db	2542	RSGTWKRHSKH-----SSSLPRVSTWRRTGSSSSILSASSESEKAKSEDEKH---	2590
Qy	2154	QLSAPLPAPLYSPGACSPVLDLRPPSDLYLPPDHCAPARGSPHSEGGKRSPENKTS	2213
Db	2591	-----VNSISGTK-----QSKENOVSAKGTWRIKENEFSPTNSTS	2626
Qy	2214	--VLGGEDGIE-----PVSPPEGMTPEGHRSRAVYPLLYRDGEQTEPSRMGSKSPGNT	2265
Db	2627	QTVSSGATNGASKTLIYMAAPVSKTEVDVVRVEDCPI-----NNPRSGRSPGTNT	2678
Qy	2266	SQPPAFPSKLTESNAWVSKKQEIKNKLNTHNRNEPYNISQPGTEIFNMPAITGTGLM	2325
Db	2679	--PP-----VIDSVSEKANPNIKDSKONQAKONY-----GNGSV	2710
Qy	2326	TYBSQAVQEHASTNMCLEAIBKALMGKYDQWEEESPPLSANAFNPLNASLPAAMPITA	2385
Db	2711	PMRTVGLNRLNSFTQVDAPOQKGTEIKPGQNNPVVSETN-----ESSIVERTPFSS	2763
Qy	2386	ADGRSDHRTLSPGGGGKARVSG---RPSRRKAKSPAGLASGDRPPSVSVHSEGDGNRR	2442
Db	2764	SSS-SKH--SSPSGTVAARVTFPNYNPSPRSSAD---STPSARPSQIP-----	2805
Qy	2443	TPUTNRWEDRPSAGST	2460
Db	2806	TPVNNNT-KRKDSKTDST	2822

RESULT 11
 US-08-452-654-7
 : Sequence 7, Application US/08452654
 : Patent No. 5691454
 : GENERAL INFORMATION:
 : APPLICANT: ALBERTSEN, HANS
 : APPLICANT: ANAND, RAKESH
 : APPLICANT: CARLSON, MARY
 : APPLICANT: GRODEN, JOANNA
 : APPLICANT: HEDGE, PHILIP J.
 : APPLICANT: JOSLYN, GEOFF
 : APPLICANT: KINZLER, KENNETH
 : APPLICANT: MARKHAM, ALEXANDER F.
 : APPLICANT: NAKAMURA, YUSUKE
 : APPLICANT: THLIVERIS, ANDREW
 : TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 : TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 : NUMBER OF SEQUENCES: 94
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Banner, Birch, McKie & Beckett
 : STREET: 1001 G Street, NW
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20001-4598
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/452,654
 : FILING DATE: 25-MAY-1995
 : CLASSIFICATION: 536
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/741,940
 : FILING DATE: 08-AUG-1991
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Kagan, Sarah A.
 : REGISTRATION NUMBER: 32,141
 : REFERENCE/DOCKET NUMBER: 1107.035574
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 202-508-9100
 : TELEFAX: 202-508-9299
 : INFORMATION FOR SEQ ID NO: 7:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 2842 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : ORIGINAL SOURCE:
 : ORGANISM: Homo sapiens
 : IMMEDIATE SOURCE:
 : CLONE: APC
 : US-08-452-654-7

Query Match 2.7%; Score 350.5; DB 1; Length 2842;
Best Local Similarity 17.8%; Pred. No. 1.1e-11;

Qy	50	A\$HLSPGSIIQP---QRRRPSLLSEFQ-----PCNERSQBELHPESHY	91
		: : : : : : : : : : :	
Db	739	ANINSPGSSLP\$LHVKQKALEAELDQAHLSETFDNTDNLSPRA\$HR\$KQRHKQSLYGDY	798
		: : : : : : : : : : : :	
Qy	92	LPELGKSEMFEISKRPRLLELPPDLPRPPLATGPAGEDLT---KDPSL---	141
		: : : : : : : : : : : :	
Db	799	VFDTRHDNRDNFNFTGNMVLVSPYLNTTVLPSSSSRGSLDSRKSDRSLERERGIG	858
		: : : : : : : : : : : :	
Qy	142	TGKLPEVPSPPHPTDEPELVVPRLSKHEELIQMDRVREITWVCQQISKLKKKOOOLE	201
		: : : : : : : : : : : :	
Db	859	LGNYHPATEN-----PGTSSKRGL-----QISTAAQAKV----ME	891
		: : : : : : : : : : : :	

Db 1492 ATSTPDGFCSSLSALSLEDEFFIQKQVLRIMPPVOENDNGNETESQPKESNENQEK 1551
Qy 916 DSAATCSA--DEVDRAEGD---KNRLSPRPSLLTPTGDPANASQOKPLDLKQLQR 969
Db 1552 EAEKIDSEKDLDDDDDDIEBIECEIISAMPT-----KSRKAKKP-----AQ 1597
Qy 970 AAAIPIQVTKVHEPPREDAAPTKPAPPAPPQONLQPEDAPQPGSS--PRGKSRSPAP 1028
Db 1598 ASKLPVPVARKSQPLPVKLLPSQ-----NRLOKQKHVSFTPGDDMPR----- 1640
Qy 1029 PADKEAFARAEAKLPGDPCCMTWGLPFPVPPREVIKASPHADPSAFYAPGHPPLPLGL 1088
Db 1641 ----VYVEG-----TPINFSTA-----TSL 1657
Qy 1089 HDTARVLPPTTISNPPLISSAKHPSVLEROIGALSQMSVQLHVYSEHAKAPVGPV 1148
Db 1658 SGL-----TIESPP-----NELAAGEGVRGAGSGEFEKRDITPT 1692
Qy 1149 TMGLPLMPDKKLAPFSGVKQBLSPRGQAGPPESLGV-----TAQASVLRGTALGSV 1203
Db 1693 -----EGRSTDEAQQGKTSVVTPELDDNKAEGDILAEICINAM 1732
Qy 1204 PGG-----SITKGIPTRVPSDSAITVR-----GSITHGTADVLVYKGTITRI 1246
Db 1733 PKGKSHKPFVKKIMDQVOQASASSAPKNQDKKKKPTSPVKPIPONTYRTRVRK- 1791
Qy 1247 IGEDSPSLDRGREDLSLPGHVIYEGKKGHVLSYEGGMSVTCCKED---GRSSSGPHE 1303
Db 1792 -NADSKNNLNAERFVSDNK-----DSKKONLKNNSKDNKLPNNEDRVGFSFADSPHH 1845
Qy 1304 TAAKRTYDMMGRVGRALISSASIEGL-----MGRAPPERHSPHLEKQHIIRGSI 1355
Db 1846 -----YTPISGTYCFSRNDSLSLDFDDDDVLSR-----EKAEELKAKENKESAKV 1894
Qy 1356 TQGIPTSVYEAQEDYLREAKLLKREGTPPP-----PPSRDL-----TEAYTKQ 1400
Db 1895 TSHTELTSNOQSANKTOATAIKAPINRGOPKPILOKQSTFPQSSKIDIPRGAATDEKLQNF 1954
Qy 1401 ALGPLKLAHGLVATVKEAGRSITHEPRELRT-----PELPLAPRLKEGSIT 1452
Db 1955 AIENTPVCFSHNSLSLSDIDDOENNNENEFIKETEPDQSGEFSKFOASAYAKPSHV 2014
Qy 1453 QGTPKYDTGASTG-----SKKHVRSRLIGSPGR 1482
Db 2015 EDTPVCFSRNSSLSDISEDDLQECISSAMPKKKPKSRLLKGDNEKHSRPMGILGE 2074
Qy 1483 TPPPVLHVDVADALERACVYESLSKSPGTASSGSGSIARGAPVIVPELCKPQRSPLT 1542
Db 2075 -----DLTLDLKDIOQPOSEHGLS--PQSENFWDKAIQOAGANSIVSSL----- 2115
Qy 1543 YEDHGAPFAGHLPR-----GSPVTWREPTPRLOEGSLSSSKASQDRK 1584
Db 2116 ---HQAAAACLSRQASDSLSLSKSGISLGSFHL---TPDQEEKPFTSNKG----- 2164
Qy 1585 LTSTPREIAKSPHSTVPEHHPHPISPYEHLLRGV--SGVDLYRSHITPLAFDPTSPRGIPL 1643
Db 2165 ----PRILKPGKSL-----ETKKIESKSGIKGKKVKS----- 2197
Qy 1644 DAAAYVLPRLHAPNTPVPHLYPPVLIIRGYPTALENQTIINDYIT'SQQMHHTATAM 1703
Db 2198 -----LITG-----KVRNSSEISGOM----- 2213
Qy 1704 AQRADMLRGLSPRESSLALNAAAGPRGIDLSQVPHLPVLPPTGTATAMDRLAYLPT 1763
Db 2214 -----KQPLQANPISRG-----RTMHIHGV----- 2236
Qy 1764 APQPFSSRHSSPLSPGGPHTLTKPTTSSSERDRDRDRDREREKSLTSTTT--- 1820
Db 2237 ----RNSSSTSPVSKKGGP-LKTPASKSPSEG-----TATTSRPG 2273
Qy 1821 -----VEHAPVWRPGTQSSSGSSSGSGSGSSSPASHAHQHSPISPRTQDALQR 1874
Db 2274 AKPSVKSELSVAVRO-TSQIGGSKAPSRSGSRDSTPRPAQQLSRPIQSPGRNSI--- 2329

Qy 1875 PSVLHNTGMKGIITAVEPSKPTVLRSTSTSSSPVRAAATFPATHCPLGCTLDGVYP---- 1930
Db 2330 -----SPGRNGI---SPPNKLSQLPRTSPSTA-----STKSSGSGKMSYTSFGROM 2373
Qy 1931 ---TLMPEVLLPKEAPRVARPERPRADTGHAFKAPPARSGLEP-----ASSPSKSGE 1980
Db 2374 SOONUTKTGSLKNASSI-----PRSESASKGLNOMNGNGANKKVELSRMSSTKSSGE 2428
Qy 1981 ----PRPLVPPVSGHATARTPA--KNLAPHASPPDPAPPASADPHREKTKQSPFSIQE 2035
Db 2429 SDRSERPVLRQSTRIKEAPSTLRKLEESASFESLSPSSRPASPTRSOAQT--PVLSPS 2487
Qy 2036 LELRSLGVHSGSYSEGEVPSVSPSLTHDKGLPKHLELDKSHLEGELRPKQPGPVK 2095
Db 2488 LPDMSLSTH--SSVQAGWKRKLPPNLSPTIEYNDGRPAKRDHARSHESPSL---PIN 2542
Qy 2096 LGG--EAAHLPHLRPLPESQSSSPLLQATAPGVKGHQRVVTLAQIHSEVITQDYTRHHQ 2153
Db 2543 RSGTWKREHSHK-----SSSLPRVSTWRTGSSSLSLSASSESSEKAKSEDEKH--- 2591
Qy 2154 QLSAPLAPLYFPGASCPLDLRRPPSDLYLPPDPHAGARGSPHSEGGKRSPPENKTS 2213
Db 2592 -----VNSISGTK-----QSKENQVSAKGTWRKIKENEFSPNTNSTS 2627
Qy 2214 --VLGGGEDGIE-----PVSPPEGTEPHGHSRAVYPLYRDGEQTEPSRMGSKSPGNT 2265
Db 2628 QTVSSGATNGAESKTLIYOMAPAVSKTEDVWVRIEDCPI-----NNPSRGRSPTGNT 2679
Qy 2266 SOPPAFFSKLTESNAMYKSKKQKQKINLTHNRNEPEYNISQPGTEIFNMPAICTGLM 2325
Db 2680 --PP-----VIDSVSEKANPNIKOSKDNQAKQNV-----GNGSV 2711
Qy 2326 TYRQAOVQEHASTNGLEALIRKALMGKYDQWEEESPPLSANAFNPLNASASLPAAMPITA 2385
Db 2712 PMRTVGLNRLNLSFIQVDAPQKTEIKPGQNNPVVSETN-----ESSIVERTPFSS 2764
Qy 2386 ADGRSDHPLTTPGGGGKAKVSG---RPSRRAKSPAPGLASGRDPPSVSVSHSEDCNRR 2442
Db 2765 SSS-SKH--SPSGTVAARVTFNPNPSRKSAD-----STSRPSQIP----- 2806
Qy 2443 TPLTNRVNEDRPSSAGST 2460
Db 2807 TPVNNNT-KKRDSTKDTST 2823

RESULT 13

US-08-452-655B-7
; Sequence 7, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk


```
Db 2075 -----DLTLDKDIQRPDSEHGSL--PDSNFWDKAIQGANSSVSSL----- 2115
Qy 1543 YEDHGAPFAHLP-----GSPVWREPTRLQEGSLSSKASQDRK 1584
Db 2116 ---HQAAACLSRQASSDSDLSLKSGISGSPHL---TPDQEEKPFTSNKG----- 2164
Qy 1585 LSTPREIAKSPHSTVEHPHPISPYEHLRGV--SGVDLYRSHIPLAPDPTSIPIRGIPL 1643
Db 2165 ---PRILKPGKSTL-----ETKIESESGIKGGKKVYS----- 2197
Qy 1644 DAAAAYLPRHLPAPNPTPHLYPPIIRGYPTAALENQTTINDYITTSQQMHHTATAM 1703
Db 2198 -----LITG-----KVRNSNEISQGM----- 2213
Qy 1704 AQADMLRGLSPRESSALNAYAGRGIDLSQVPHLPVLVPTPTGTPATAMDLAYLPT 1763
Db 2214 -----KQPLQNMPSISRG-----RTMIHIPGV----- 2236
Qy 1764 APQPFSSRHSSPLSPGGPHTLTKPTTTSSSRERDRDRDRDREREKSILTSTTT--- 1820
Db 2237 ---RNSSSSTSPVSKGPP-LKTPASKSPSEQ-----TATTSRPG 2273
Qy 1821 -----VEHAPTWRCGTQSSSGSSGGSSGSSSRPASHAHQHPISPRTDALQOR 1874
Db 2274 AKPSVKSELSPVARQ--TSQIGGSKAPSRSGSRDSTPSRPAQOPLSRPTQSPGRNSI--- 2329
Qy 1875 PVLHNTGMKGITAVEPSKPTVLSTSTSPVPAATPPATHCPGLGTLGVYP--- 1930
Db 2330 -----SPGRNGI---SPPNKLSQLPTSPSTA-----STKSSGGSKMYSYTFGRQM 2373
Qy 1931 ---TLMEPVLLKPEAPRVARPERPRADTGHAFLAKPPARSGLEP-----ASSPSKSGE 1980
Db 2374 SOONLTQKGLSKNASSI-----PRSESASKGLQNMNNGANKKVELSRMSSTKSSGSE 2428
Qy 1981 ---PRPLVPVSGHATARTPA--KNLAPHASPPPPAPPASDPHREKTKOSKPFSTQE 2035
Db 2429 SDRSERVLVROSTTIKEAPSTLRKLEESASFESLSPSPASPTRSQAOT--PVLSPS 2487
Qy 2036 LELRSLGYHGSSYEGVEPVSPVSPSLTHDKLPKHEELDOKHLEGLERPKQPGPVK 2095
Db 2488 LPDMSLSTH--SSVQAGGWRKLPNLSPTIEYNDGRPAKRHDIAHSHSEPSRL-----PIN 2542
Qy 2096 LGG--EAAHLPHRLPESQSSPLLQATPGVKGHQVRVYLAQHISEVITODYTRHHPO 2153
Db 2543 RSGTWKREHSHK-----SSSLPRVSTWRTGSSSILSASSESSEKAKSEDEKH--- 2591
Qy 2154 QLSAPLPAPLYFPGASCPLVLRPPDLPLPPDHGAPARGSPHSEGGKSPFPNKTS 2213
Db 2592 -----VNSISGTK-----QSKENQVSAKGTWRKIKENEFSTPNSTS 2627
Qy 2214 --VLGGGEGDIE-----PVSPPEGMTEFGHSRAVYPLLYRDGEQTEPSRMGSKSPGNT 2265
Db 2628 QTVSSGATNGAESKTLIYQMAFAVSKTEDVWVRIEDCPI-----NNPRSGRSPTGNT 2679
Qy 2266 SPPAFFSKLTESNAMYKSKQEIKNKLNTNHRNEPEYNISQPGTEIFNMPAITGTGLM 2325
Db 2680 --PP-----VIDSVSEKANPNKDKDNQAKQNV-----GNGSV 2711
Qy 2326 TYRQAVQEHASTNGLAEIRKALMGKYDQWEEFPPLSANAFNPLNASALPAAMPITA 2385
Db 2712 PMRTVGLNRLNSITQVDAPDQKGTETKPGQNNPVVSETN-----ESSIVERTPFSS 2764
Qy 2386 ADGRSDHTLTPSGGGKAKVSG---RPSRRKAKSPAPGLAGDRPPSVSVHSEDCNR 2442
Db 2765 SSS--SKH--SSPSGVVAARVTPFNTNPNRPSKSSD-----STSRPSQIP----- 2806
Qy 2443 TPLTRVREDPSSAGST 2460
Db 2807 TPVNNNT--KKRDKTDST 2823
```

RESULT 14

US-08-450-582-2

```
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-450-582-2
```

```
Query Match 2.7%; Score 350.5; DB 3; Length 2843;
Best Local Similarity 17.8%; Pred. No. 1.1e-11;
Matches 473; Conservative 357; Mismatches 1007; Indels 821; Gaps 109;

Qy 50 ASHLSPPGSIQIP--QRRRPSLLSEFQ-----PGNERSQELHLRPESHY 91
Db 740 ANINSPGSSLSLHVYRKQKALELDAQHLSETFDNIDNLSPKASHRSKQRHKQSLGY 799
Qy 92 LPELKSMEWETESKRPLELLPDLLRPSPLATGQAGSEDLT---KDRSL----- 141
Db 800 VFDTRHDDNRSDNFNTGMTVLSPLYLNTVLPSSSSSRGSLDSSRSKDRSLERERG 859
Qy 142 TGKLEPPSPPPHTDPELELVPPRLSKEELIQNDVRDREITWVEQIISKLKKQOOLE 201
Db 860 LGNYHPATEN-----PGTSSKRGLE-----QISTTAQAIKV-----ME 892
Qy 202 EEAAPPEPEKVPSPPTTESKHSRLVQIYDNR--KKAEAHRILEGIGPQVELPLYNQ 259
```


QY 866 CTEAEEG-----PAKGDAAEAATAEGALKAKEGSGGRATTA---KSSGAPQDS 915
Db 1492 ATESPDPGSCSSLSALSALDPEFTKQVELKIMPPVQENDNGNETESQPKRESNEQBK 1551
QY 916 DSSATCSA--DEVDEAEGD---KNRLLSPRSLTPTGDPANASQKPLDLKQKOR 969
Db 1552 EAEKTDSEKDLDDDDDDIEILEECIISAMPT-----KSSRAKAKP-----AQ 1597
QY 970 AAAPQIOWTVKHEPPREDAAATKAPAPPQPPONLOPESDAPQPGSS--PRGKSRSAP 1028
Db 1598 ASKLPVPVARKPSQLPVYKLLPSQ-----NRLOPKHVSFTPGDDMP----- 1640
QY 1029 PADKEAFAAEAKLPGDPPCMTSGLPFPVPPREVIKASPHADPSAFVAPGCHPLPLGL 1088
Db 1641 -----VYCVG-----TPINFSTA-----TSL 1657
QY 1089 HDTARVLP RPPTINPPPLISSAKHPVLERQIGAISOGMSVQLHVPYSEHAKAPVGV 1148
Db 1658 SDL-----TIESPP-----NELAAGEVARGAQSGEFEKRDPTIPT 1692
QY 1149 TWGLPLPMDPKKLAFSGVKQBLSPRGOAGPPESLQVP-----TAQEAASVLRGTALGSV 1203
Db 1693 -----EGRSTDEAQQGKTSSVTPELDDNKAEEGDILAECINSAM 1732
QY 1204 PGG-----SITKGIPSTRVPSDSAITYR-----GSITHGTADVLYKGTITRI 1246
Db 1733 PKGSHKPFVVKKIMDVQOASASSAPKNQLDGKKKPTSPVKPIQONTYRTRVRK- 1791
QY 1247 IGEDSPRLDRGREDSPKGVHYEGKGHLYSYEGGMSVTCCKED-----GRSSGPPHE 1303
Db 1792 -NADSKNNLNAERFVSDNK-----DSKQNLKNSKDFNDKLPNNEDVRGSAFADSPHH 1845
QY 1304 TAAPKRTVDMGVRGVRALSSASIEGL-----MGRAPPERSPHILKEQHHRGSI 1355
Db 1846 -----TPIEGTYCFSRNDSLSLDFDDDDVLSR-----EKAELRAKENKESEAKV 1894
QY 1356 TOGIPRSTVEAEDYLRREALKKREGTPPP-----PPPSRDL-----TEAYKTQ 1400
Db 1895 TSHTELTSNQOSANKTOAIKQPINRGPKPILOKQSTFPQSSKDIPDRGAATDEKLQNF 1954
QY 1401 ALGPILKPAHGLVATVKEAGRSTHETPRELRUT-----PELPLAPRLKEGST 1452
Db 1955 AIENTPVCFSHNSLSLSDIDQENNNKENEPKETEPDQSGEPSPQOASYPAPKSEHV 2014
QY 1453 QGTPLKYDTGASTTG-----SKHDRVSLIGSPGR 1482
Db 2015 EDTPVCFSRNSSLSDSEDDLLQECISSAMPKKKPKSRLLKGDNEKHSRNMGGILGE 2074
QY 1483 TFPVPHPLDVMDARALERACYEESLKRPGTASSGGSIGARGAPVIVPELGPQSPILT 1542
Db 2075 -----DLTLDLKDQRPDSEHGLS--PDSENFDRKAIQEGANSIVSSL----- 2115
QY 1543 YEDHGAPFAGHLPR-----GSPVTMRPTPRLOEGLSSSSKASQDRK 1584
Db 2116 ---HQAACLSRQASDSDSILSKSGISLSPFHL-----TPDOEKPFSTNKG----- 2164
QY 1585 LTSTPREIAKSPHVPHEHPHPISPYEHLRGV--SGVDLYRSHIPLAFDPTSIPIRGILP 1643
Db 2165 ----PRIKPGEKSTL-----ETKKISESGIKGGKGVKVS----- 2197
QY 1644 DAAAAAYLPHLAPNTYPHYLPYPPYLPDPTAALENRQTIINDYITTSQOMHNTATAM 1703
Db 2198 -----LITG-----KVRNSEIISQOM----- 2213
QY 1704 AORADMLRGLSPRESSLALNWAAGRGIIDISQVPHLPVLVPPPTGCTATAMDRLAYLPT 1763
Db 2214 -----KQPLQANPESISRG-----RTMTHIPGV----- 2236
QY 1764 APQPFSSRRHSSPLSPGGFTHLTKTTTTSSSERDRDRDRDREREKSIILTSITT--- 1820
Db 2237 -----RNSSSSTSPVSKGPP-LKTPASKSPSEGO-----TATTSPRG 2273

QY 1821 -----VEHAPIWRPCTEQSSSGSSGGGGSSSRPASHAHQSPISPRTOTDALQOR 1874
Db 2274 AKPSYKSELSPVARQ--TSOIGSSSRAPSGSRDSTPSRPAQOPLSRPTQSPGRNSI--- 2329
QY 1875 PSVLHNTMGKGIITAVEPSKPTVLRSTSTSSPVSPAATFPFATHCPHGLGTLGVYP--- 1930
Db 2330 -----SPGRNGI---SPPNKLSQLPRTSPSTA-----STKSSGSKMSYTSPOROM 2373
QY 1931 ---TLMEPVLLKEAPRVARPERPRADTGHAFKLAKPARSGLEP-----ASSPKGSE 1980
Db 2374 SOONLTQGLSKNASSI-----PRSESASKGLNOMNNGANKKVELSRMSTKSSGSE 2428
QY 1981 ----RPPLVPVSGHATIAITPA--KNLAPHASPODPAPASADPHRKTQSKPFSIOE 2035
Db 2429 SDRSERPVLVROSTFTKEAPSTLRKLEESASFESLSRSSRSPASPTRSOAQT--PVLSPS 2487
QY 2036 LEIRSLGYHGSYSRPGVEPVSPVSPSLTHDKGLPKHLELDKSHLEGELRKPQGPVK 2095
Db 2488 LPDMSLSTH--SSVQAGGWKRLPPLNLSPTTEYNDGRPAKRHDIAARSHSESPSL---PIN 2542
QY 2096 LGG--EAAHLPHLRPLPESQSSSPLLOTPAGVKGHORVVTLAQHLSEVITQDYYRHHQ 2153
Db 2543 RSGTWKREHSHK-----SSSLPKVSTWRRTGSSSSILSASSESEKAKSEDEKH--- 2591
QY 2154 QLSAPLPAPLYSPFGASCPCVLDLRRPPSDLYLPPDPHGAPARGSPHSEGGKRSPEPNKTS 2213
Db 2592 -----VNSISGTK-----QSKENQVSAKGTWRKIKENEFSPNTS 2627
QY 2214 ---VLGGEDGIE-----PVSPEGMTEPGHRSRASYVPLLYRDGEQTEPSRMGSKSPGNT 2265
Db 2628 QTVSSGATNGAESKTLIYQMAPAVSKTEDVWVRIEDCPI-----NNPRSGRSPTGNT 2679
QY 2266 SOPPAFFSKLTESNAMYKSKQETNKKLINTHNRNEPEVYNISOPGTEIFNMPAITCTGLM 2325
Db 2680 ---PP-----VIDSVSEKANPNKDKDNQAKQNV-----NGSV 2711
QY 2326 TYRQAOVQEHASTNMGLEAIIRKALMGKYDQWEEESPPLSANAFNPLNASASLPAAMPITA 2385
Db 2712 PMRTVGLNRLNSFIQVDAPDQKGTEIKPGQNNPVVSETN-----ESSIVERTPFSS 2764
QY 2386 AGRSDHLLTSGGGGKAKVSG---RPSRKAASAPAGLASGDRPPSPSVSVHSEGCNRR 2442
Db 2765 SSS-SKH--SSPSGTVAARVTFPNYNSPRKSSAD---STSAARSPQIP----- 2806
QY 2443 TPLTNRWEDRPSAGST 2460
Db 2807 TPVNNNT--KKRDSKTDST 2823

Search completed: September 8, 2001, 14:35:09
Job time: 17211 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 09:57:13 ; Search time 70.16 seconds
(without alignments)
2732.773 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQTWRATEPRYP.....WDEPKPLCSQYETLSDSE 2517

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7871	59.6	1495	2 S60255	transcription co-r
2	4187	31.7	2453	2 S60254	nuclear receptor c
3	1272	9.6	1047	2 T46489	hypothetical prote
4	574	4.3	5327	2 T13564	microtubule-associ
5	552	4.2	2649	2 T51023	hypothetical prote
6	526	4.0	2142	2 B35098	MHC class III hist
7	520	3.9	2187	2 T30826	nascent polypeptid
8	512	3.9	5762	2 A41819	proline-rich pepti
9	508.5	3.8	1872	2 S36152	MHC class III hist
10	504	3.8	3942	2 T42730	Bassoon protein -
11	503.5	3.8	7962	2 T38346	elastic titin - hu
12	503	3.8	1870	2 S37671	MHC class III hist
13	502.5	3.8	2715	2 T13049	eyelid - fruit fly
14	498	3.8	3938	2 T42761	Bassoon protein -
15	492.5	3.7	2774	2 A43359	microtubule-associ
16	477	3.6	2282	2 T42717	DNA-binding protei
17	470.5	3.6	4957	2 T03455	ALR protein - huma
18	470.5	3.6	5262	2 T03454	ALR protein - huma
19	445.5	3.4	1952	2 T48814	hypothetical prote
20	436	3.3	1151	2 T18535	high molecular mas
21	424.5	3.2	2464	1 QRASPI	microtubule-associ
22	419	3.2	5170	2 T15348	hypothetical prote
23	416.5	3.2	1791	2 T02345	hypothetical prote
24	412.5	3.1	1226	2 S15053	hypothetical prote
25	407.5	3.1	1464	1 CGHUI5	collagen alpha 1(I
26	406.5	3.1	2364	2 A56577	microtubule-associ
27	404	3.1	1274	2 T16251	hypothetical prote
28	403	3.0	1110	2 I51116	NF-180 - sea lamp
29	397	3.0	1670	1 CGHU3B	collagen alpha 3(I

30	395	3.0	1520	2 T00273	hypothetical prote
31	394	3.0	2783	1 A41948	alpha-fetoprotein
32	391	3.0	3924	2 S37431	ankyrin 2, neurona
33	389.5	2.9	1188	2 S49315	extensin-like prot
34	389.5	2.9	3869	2 A48205	All-1 protein +GRE
35	389	2.9	1320	2 JC5630	TCOF1 protein - mo
36	383.5	2.9	3968	2 A44265	trithorax homolog
37	383	2.9	2944	2 A54849	collagen alpha 1(V
38	377	2.9	1691	1 S22917	collagen alpha 5(I
39	375	2.8	1006	2 T42731	atrophin-1 related
40	374.5	2.8	2722	2 T20532	hypothetical prote
41	373.5	2.8	1184	2 S50832	atrophin-1 - human
42	373.5	2.8	1494	2 T14355	protein-tyrosine-p
43	373.5	2.8	2938	2 T30249	cell proliferation
44	373	2.8	1763	2 S16366	collagen alpha 2(I
45	372.5	2.8	1173	2 T31421	C-terminal domain-

ALIGNMENTS

RESULT 1

S60255
transcription co-repressor SMRT - human
C:Species: Homo sapiens (man)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 05-Nov-1999
C:Accession: S60255
R:Chen, J.D.; Evans, R.M.
Nature 377, 454-457, 1995
A:Title: A transcriptional co-repressor that interacts with nuclear hormone receptors
A:Reference number: S60255; MUID:96008552
A:Accession: S60255
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1495 <CHE>
A:Cross-references: EMBL:U37146; NID:g1045654; PIDN:AAC50236.1; PID:g1045655

Query Match 59.6%; Score 7871; DB 2; Length 1495;
Best Local Similarity 100.0%; Pred. No. 3.1e-269;
Matches 1487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1031	DKEAFAAEAQKLGDP	PCWTSGLPFPVPPREV	IKASPHADPPSAFSAF	YAPGHPGLPLGLHD	1090
DB	9	DKEAFAAEAQKLGDP	PCWTSGLPFPVPPREV	IKASPHADPPSAFSAF	YAPGHPGLPLGLHD	68
QY	1091	TARPVLPRPTTISNPP	PLISSAKHPSVLERQIC	AIQSGMSVQLHVPYSE	HAKAPVGPVTM	1150
DB	69	TARPVLPRPTTISNPP	PLISSAKHPSVLERQIC	AIQSGMSVQLHVPYSE	HAKAPVGPVTM	128
QY	1151	GLPLMDPKKLAPFSG	VKQOLSPRQAGPPESIG	VPQAQESVLRGTALGS	VPGGSITK	1210
DB	129	GLPLMDPKKLAPFSG	VKQOLSPRQAGPPESIG	VPQAQESVLRGTALGS	VPGGSITK	188
QY	1211	GIPSTRVPSDSAITYR	SGTTHGTPADVLKGTIT	RIIGEDSPSRDRGRED	SLPKGHVY	1270
DB	189	GIPSTRVPSDSAITYR	SGTTHGTPADVLKGTIT	RIIGEDSPSRDRGRED	SLPKGHVY	248
QY	1271	EKKKGHVLSYEGGMS	VTCQSKEDGRSSSGPP	HETAPKRTYDMMEGR	VGRAISASIEGL	1330
DB	249	EKKKGHVLSYEGGMS	VTCQSKEDGRSSSGPP	HETAPKRTYDMMEGR	VGRAISASIEGL	308
QY	1331	MGRAPIPRHSPHLKEQ	HHIRGSITQGIIPRSV	EAQDYLRREKLLKREG	PPPPPPS	1390
DB	309	MGRAPIPRHSPHLKEQ	HHIRGSITQGIIPRSV	EAQDYLRREKLLKREG	PPPPPPS	368
QY	1391	RLTEAYKTAQALGPL	KLKPAHEGLVATVKEA	GRSIIHEIPREELRHT	PELPAPRLKEGS	1450
DB	369	RLTEAYKTAQALGPL	KLKPAHEGLVATVKEA	GRSIIHEIPREELRHT	PELPAPRLKEGS	428
QY	1451	ITQGTPLKYDTGAST	TGSKKHDVRSLLIGSP	GRTPFPVPHLDVNADA	RALRACYEESLKS	1510
DB	429	ITQGTPLKYDTGAST	TGSKKHDVRSLLIGSP	GRTPFPVPHLDVNADA	RALRACYEESLKS	488

QY 829 EE-----TAAAPVVERBEG-----KPPAAELAVDTGKAEPVVKSECTEAE 871
Db 795 VEAQVTDASAETAEPMDVDHECGAEGSSVLDPPAPTK--ADSVDPENQV--PENTASKG 851
QY 872 EGPAKGDAEAAEAATGALKAKEKGGSGRATTAKSSGAPO---DSDSATCSADEVIDE 928
Db 852 EDGARERDLEST-----SEKTEARDEDVVVAEQIERPEPQSDDDSDATCSADEGYD 902
QY 929 AEGGDKNRL--LSRPSLLTPTGDPANASPOKP--LDLKOLKQRAAAATPPI----- 976
Db 903 GE--PERQVFPMDAKPSLLTPPGSILI--SSPIKNLLDLPLQOHRAAVIPPMVWCTPCNI 960
QY 977 -----QVTKVHEPPREDAAPTKPPAPPQQNQLPESDAPQPGSSPRGK 1022
Db 961 PIGTPVSGYALYQRIKAMHESALLE-----EQRQREQVDLECRSSTSPCST 1008
QY 1023 SRSPAPPADKEAFAEQAOKLPCDPCWTSGLFPFPPPREVIKASHAPDPSAFSAPGCH 1082
Db 1009 SKSP-----NRE-----W-----EVLQAPAP-----QVITNLPEGV 1034
QY 1083 PLPLGLHDTARVLPPTISNPPPLISSAKHPSVLERQIGAISOG-----MSVOLHVPY 1137
Db 1035 RLP-----TTRTRPPPLIPSSKTTVASEK--PSFI--MGGSISOGTPTGLSSHNQAYP 1086
QY 1138 SEHAKAPGVMTGLPMDPKKLPFGSGVQEOQLSPRQAGPPESLSGYPTAQEAASVLRG 1197
Db 1087 QEAPKPSVGSISLGPLRQOESTKAAPLTYIKQEEFSPRSQNSQPSGLLV--RAQHEGVVRG 1145
QY 1198 TALGSPVSGSTKGLPSTRVPSDSAITVRSITHGTGA-----DVLKGTITRI--IG 1248
Db 1146 TA-GAVQSGSTRTPAKSAISVETISSRLSGITQGTTPALPQAGIPTALVKGVPVSRMPTE 1204
QY 1249 EDSPLRDGRSDLPKGVHVIYEGKGVHLSVYEGGVSMTQCKEDGRSSGPPHETAAPK 1308
Db 1205 ESSPEKV--REEAASKGVHVIYEGKSGHLSYDNTKNA-----REGTRSPRTAHMSL--K 1255
QY 1309 RYDMMEGRVGRATIS-----SASIEGLMRAIPERHSPH--HLKEQHRIIGSTIQGIPR 1361
Db 1256 RSYEAEGSIKQGMSPSVSAPEGLICRALP--RGSPHSLDKERTVLSGSIQMGTPR 1313
QY 1362 SVVEAQEDYLREAKLLKREGTPPPPPSRDLTEYAKTQALGPLKLPKAHEGLVATVKEA 1421
Db 1314 ATAESFEDGL--KYPQKIKRES-----PPIRAFEGAI-----TKGKP--YDG--ITTIKEM 1358
QY 1422 GRSIHEIPREEL-----RHTPELPLAPRLKEGSTTQGTPLKYDYGATTSKKHDEVS 1476
Db 1359 GRSIHEIPRODILQESKRTPEVQVOSTPIIEGSIQGTPIKFDNN--SGQSAIKHNKSL 1417
QY 1477 IGSPGRTPPVPHLDVMD--ARALERACYE-----ESLKSRTPGTASSSGGSTARGAPVIV 1530
Db 1418 ITGPSKL--PRGMLFIVPENIKVVERGYEDVKAGEPYRARTSVVSSGSPVLRST---L 1472
QY 1531 PELGKPROSPUTYEDHGA-----PFAGHLRPGSPVTMREPTPRLOEGLSSSK--ASODRK 1584
Db 1473 HEAPKAQLSPGLYDDSSARRTFVSYQNTISRGSPMNNR-----TSDVSSSKSASHERK 1525
QY 1585 LSTSTPRE-----IAKSPHSTVPEHHPHTISPYEHLRLGVSGVDLYRESHIPLAFDPTSIPIRG 1640
Db 1526 STLPTQRESITPAKSPVGVDPDIYSH--SPDPDHRSSAAAGEVYRSHLPLHLDP--AMPFH 1582
QY 1641 IPLDAAAAYLPHRLAPNPTYPHLYPYLIRGYPDTALEN--RQTIINDYITTSQOMHNT 1699
Db 1583 RALDPAAYALLQRLSPPTPGPSQYLY-----AMENRTQILNDYITTSQOMQVNL 1633
QY 1700 ATAMAQRADMLRGLSPRESSLALNYAGPRGIIDLSQVPHLPVLPVPPPGTATAMDRLA 1759
Db 1634 -----RPDVTGRGLSPREQLGLPYPA--TRGIIDLTNNPP--TILVPHAGGTSTPPMDRIT 1685
QY 1760 YLPTAPQPFESSR--HSSSPLSPGGTHLTKPTTSSSERERDRDRDRDREREKSILTST 1818
Db 1686 YIPGTQVTPPFPYNAASLSGPHTHL-----AAAASAEEREREREKERERERERER 1742

QY 1819 TTVEHAP-----IWRPGTEQSSSGSGSGSGSSRSPASHASHAHQHSPIRPTQDA--LQOR 1874
Db 1743 ERIAAPADLYLRPGSEQ-----PGRPGSHGYVRSPP--SVRTQETILQOR 1787
QY 1875 PSYLHNTGKGLITAVEPSKPTVLSTSTSSPV-----RPAATFPATHCPGLGTLGVYVP 1930
Db 1788 PSVFOGTNGTSVITPLDPTAQLRIMPLPSGGPSISQGLPASRYNTAADA--LAALVDAAAS 1846
QY 1931 TLMPEVLLPKAPR-----VAPRPRADTGHAFKAPPARSGLEPASSPK 1977
Db 1847 AQOMDVSKTESKHEAARLEENLRSAAVSEQOQLEKNEVEKRSVOCVCTSSALPSG 1906
QY 1978 GSEPRPLV-----PPVSGHATARTPAKN--LAPHHASPDPPAPPASASDPHRE 2024
Db 1907 KAQPHASVYVSEAGKDKGPPKSRVEEELTRGKITTAANFIDVITITQIASDKDARER 1966
QY 2025 KTQSPFSTQLELSLGHSGSYSPGVEPVSPVSSPSLTHDKGLPKHLEELDKSHLSE 2084
Db 1967 GSQSSDSS-----SSLSSHYETASDAIEVISPASSPAPPOEKPQAYQPMVKANQAE 2020
QY 2085 ELRPKQPGPVKLGGEAAHLPLR-----PLPESQSSSSPLLQ--APGVKQHQRVVTLAQ 2137
Db 2021 ESTROYEGP-----LHYSRQOESPQQOPLPSSQSEGMQVPRTHRLITLAD 2071
QY 2138 HISEVITQDYTRHH--PQQLSAPL-----PAPLYSFPGASCPCVLDLRRPPSDLYLPPPD--- 2189
Db 2072 HICOLIITQDFARNQVPSQASTSTFOTSPALSSTP-----VTKTSSRYSPESQST 2123
QY 2190 --HGAPA--RGSHP-----SEGGKRSPEPNKTSVLGGGEGDIEPVSPPEGMTEPG--HSRSA 2240
Db 2124 VILHPRPGPRVSPENLVDSRGRPGKSPERSHI---PSEPYEIPSPQG---PAVHEKQD 2177
QY 2241 VYPLLYRQGEQTEPSRMGSKSPGNTSQPAPFESKLTESAMVSKKQKQINKLTHNRN 2300
Db 2178 SMLLSQRGVDPADORSRSPGISYILPFFTKL--ESTSPMVSKKQEIFRKLNSGGG 2236
QY 2301 EPEYNISOPGTEIFNMPAITGTGLMTYRSQAVQEHASTNMGLEAIIRKALMGKYQWEE- 2359
Db 2237 DSDMAAAGPTEIFNLPAVTTSGAVSSRSHSPADPAS--NLGLIEDIRKALMGSPDDKVED 2295
QY 2360 -----SPPLSANAFNPLNASASLPAAMPITAADGRSDHITLTPSGGG--GKAKVSGRPSRK 2413
Db 2296 HGVNMSHEV--GIMPGSASTSV-----VTSSEARRDEGEPSHAGVCKPKLINKSNRK 2347
QY 2414 AKSPAPGLA--SGDRPPSVSVHSECDNRRPTLNRVREDRPSAGSTPPFPYNPLIMRL 2471
Db 2348 SKSPIPGOSYLGTERPSSVSVHSEGDYHRQTP--GWAEEDRPSSTGSTQFFYNPLITRM 2405
QY 2472 QAGVNASPPPLPGLPAGSGPL--AGPH---HAMDEEPKPLLCQYETLSDSE 2517
Db 2406 ---LSSTPTQIACAPSAITQAAPHQONRIWEREPAPLLSQAQYETLSDSD 2452

RESULT 3

T46489

hypothetical protein DKFZp434M075.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T46489

R:Duisterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, January 2000

A:Reference number: 223035

A:Accession: T46489

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1047 <AAA>

A:Cross-references: EMBL:AL137641

A:Experimental source: adult testis; clone DKFZp434M075

C:Genetics:

A:Note: DKFZp434M075.1

Query Match 9.6%; Score 1272; DB 2; Length 1047;

QY 1685 IINDYITSOQHHNTATANAQADMLRLGLSPRESSIAL---NYAAGPRGIIDLSQVPHLP 1741
Db 1214 GSNVGMEDGERPRRRHGRAQODK---PPFRRLKQERENARGSEG----- 1258
QY 1742 VLVPTPTGTPATA---MDRLAYLPTAPQFSSRSHSSPLSPGPPHLTKPTTSS----- 1793
Db 1259 ---KPSLTLPASAPGPEALTITVAPAPRAAAKSPDLNSQNSQANQEWETASESSDF 1315
QY 1794 -SERERDRDRE-----RDRERERKSILTSITT 1820
Db 1316 TSERGDKKAPPVLLTPKAVGTGGGGGAVPGISAMSRGDLQRAKDLKRSFSSORP 1375
QY 1821 VEHAPITWRPGTBSQSSSSSGGGG-----SSSRPASHASHAQ 1859
Db 1376 GMERQNRPPGGKAGSGSSSGGGGPGGTGRCGDKRWPSPKNSRPERPGL 1435
QY 1860 HSPISPTODALQQRPSVLHNTGMKIITAVBPSKPTVLRSTSTSPVNRPAATFFPATHC 1919
Db 1436 PLPPPPPPSSSAVERLDQVTH-----SNPAGIQALQALSSRQGSVTAPGGH- 1481
QY 1920 PLGGITLDGYVTLMEPVLLPKAPRVARPERPRADTGHAFKAPAR-----SGLEPA 1972
Db 1482 -----PR-HKGPPOAPQGPS--PRPTRYEPQRVNSGL--- 1512
QY 1973 SSPSKGSEPRPLVPVSG-----HATARTPAK-----NLAPHAS----- 2008
Db 1513 SSDPHEEPGMVGVGTGTPRDSAGVSPFPKRRRPPKPELLQESLPPPHSSGFLGS 1572
QY 2009 -PDPAPPASAD-----PH-----REKTSKPFISQ-----ELE 2037
Db 1573 KPEGPGQAESRDTGTEALTHIWNRLHATSRKSYRPTSMPEWMEPLSPFEDVAGTEMS 1632
QY 2038 LRSGLYHGSYSYEGVEPVSPVSSSLTHDKGL-----PKH-----LELDKSHLEG 2084
Db 1633 QSDSGVDLSDGVSQSGPCSQRSSP-----DGLKGAAEGPPKRGSSPLNAVPCGGPG 1688
QY 2085 ELRPKQPGVKLGCGAAHLPHLRPLPESQSPSSPLLOTAPGVKGHVTLAQHISEVIT 2144
Db 1689 SEPPRRPPAPHDGRKELPREQPLP-----PGPIGYER---SOR-----T 1726
QY 2145 QYTRHHPOOLSAPLPAPLYSPFGACPVLDLR-----PPS-DL 2183
Db 1727 DRGTEFGPLRPS--HRPGPPVQV--GTSKDSLDRLVVGDSLKAEKELTASVTEAIPVSRDW 1784
QY 2184 YLPPPDHGAARGSPHS---EGGKRSRPN-----KTSVLGGGEDGI- 2222
Db 1785 ELLP---SAAASAEFQSKNLDGCHVPEPSSSGQRLYPEVFTVGSAGPSSQISGSGHGLS 1841
QY 2223 -----EPVSPPEGMTEPGHSRAV-YPLLYRDGEQTEPFSRMGSKS 2261
Db 1842 ITSQWRLRPGTSLHPIYRSQPLYLPPGPAPPSALLSGVALKGQFLDFSTMQATELGKLP 1901
QY 2262 PONTSOPPAFF-----SKTESNSAMVSKKQEIKNKLNTHNRNEPEYNISOPGTEI 2313
Db 1902 AGGVLYPPPSFLYSPAFCSPLPDTSLQVRO-----DLPSF-SDF 1941
QY 2314 FNMPAITG--TCLMTRYSOAVOEHAHSTNMGLEAIIIRKALMGKYDQWESPELPSANAFNPL 2371
Db 1942 YSTPLQPGQSGFLSGAPAQO-----MLPMDVDSQLPVNFGSLPPAPP---PAPPL 1992
QY 2372 NASASLPAAMPITAADGRSDHTLTSPGGGKAKVSGRFSRRKAKSPAPCLAGSDRPPSVS 2431
Db 1993 SLLPVGPALQPPSLA-----VRPPPAATRVLPSPA-----RPFPA 2029
QY 2432 SVHSE-----GDCNRRTPLNRRWEDRPPSAGSTPPFPYNPLI--M 2469
Db 2030 LGRAELHVPVELKPFODYQKLSNLGCGSSRTPPTGTRFSGLNSRLKATPSTYSGVFRQ 2089
QY 2470 RIQAGVMAAP-----PPGLPAGSGPLAGPHAWDEEP 2502
Db 2090 RVDLYQASPPDALARWIKPWERTGPPREGFSRR-AEEP 2128

RESULT 7
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:Accession: T30826
R:Yotov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing-in of a proline-rich exon converts alphanAC into a mus
A:Reference number: Z20889; MUID:96312450
A:Accession: T30826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: EMBL:U48363; NID:gl666688; PID:gl666689; PIDN:AAB18732.1
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A:Note: differential splicing converts alphanAC into a tissue-specific DNA-binding ac
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 3.9%; Score 520; DB 2; Length 2187;
Best Local Similarity 19.7%; Pred. No. 4.4e-11;
Matches 448; Conservative 219; Mismatches 782; Indels 830; Gaps 93;

QY 738 GPATVNNSDTESIPSPTHEAA-----KDTGQNGPKPPATLGADGP---PGPPT 784
Db 3 GEATETVPATEQELPOQAETAFLPMSSALKVAAVGQGPPTPPSSSLGPOQSIVTAHQPS 62
QY 785 P-PRRTSRAPTE-----TTPASEA---TGATTP-----PP----- 810
Db 63 PLPSSVSTPPEVPPAQPITAEATLPSGTAPPTPTFLPHLIGPPISPAALALASPMIGLA 122
QY 811 -----APSPSAPPVVVPKKEKEETAAAPVE----- 838
Db 123 QKGARSSAPLSLVALAPHSVQKSSVCPPHPLTSPSAAAGELGALTASIPLEPKTSTS 182
QY 839 -----EGEEKQPP-----AAEELAVDTGKAEPEPVKSE 865
Db 183 QVPSQGTNLNLAGTAPCPDVVVRAPPSHLENPLASVQPGMLSCPQLSNTSPVKGVPISSA 242
QY 866 CTEEAEEGPAKGDAAEAETAEGALKAE-----KKEGSGRAT 904
Db 243 LTQSRLSLNLGKPVSPPARNTAAPSILAPSTSLGCHLPLLHHSSVDSPIQPPGSGGLAV 302
QY 905 T-----AKSSGAPQDSDSATSCSADEVDEAEGGDK- 934
Db 303 SNPTSVGHSGIAASCPCPVCVVPALPSRLLAVDSCAAPDDKGSASVTNELCSPPGSSNV 362
QY 935 -NRLLSRPSLLTPTGDPNANASPOKPLDLKOLKORAAAIPPIQVTKVHEPP----- 985
Db 363 AGTSLSPKASLV-----PKGSNVALQPL-----VTQVPASOKTGLKEIPVSCIGATH 409
QY 986 ---REDAPTKPAAPPAPPPQNLQPESDAPQPGSSPRGKSRSPAPPADKAEFAAEAKQ 1041
Db 410 HALDNPSAISVANATHVPPPT-----SSGLVSKDPSASVTSILVPPAAHKQ 455
QY 1042 LPGDPCCWTSGLPF--PVPPREVIKASHPADPSAFSPAPPGHPLPLGL-----HDTARVPL 1096
Db 456 FPAPPASATLGVVPSPLPATEGLK---NLPIALSALVNVGAPVSPAQAGLPTKRKDTLQPLA 512
QY 1097 P-----RPPTISNPPPLISAKHPSVLERIGIATISQMSVOL 1133
Db 513 PIALKESPSOSASLEVLSEDTVTTKTGGPAPVVRPA-----IAGVATTTSIRA 563
QY 1134 HVPYSE-HAKAPVGVVTMGLPL-----PMDPKKLAPFSGVKQEQQLSPRGQAGPPE 1182
Db 564 DSPPAIVRADSCVSNPTVSQPLKRSVTDPAAMAPRTAKNTAPST--TSPLVPLASEGCPV 620
QY 1183 SLGVPTAQEASVLRGTALGVSFPGSGITKGP-----STRVPSDAITYRGS 1229

Db 621 ASSMALSPQNASVETALALSP--EIPKSVPPDPPLAEISFNARKYDAVSHMESSGS 678
QY 1230 THGTP-ADVLYKGTTRIIGEDSPRLDRGRDS-----LPGHVI 1269
Db 679 ROGHDPDASTAGTVVCL-----ADSLDTSVSKGALSASSPLYPLEVSELPFAGLA 734
QY 1270 YGKKGHVLVSIEGGMVTOCKEDG--RSSGPPHETAA----- 1306
Db 735 VOGPKGSL-----NKLSPTPSSKAGVPVSTGAPPPKGAIPVPTESSISKQVPAEILPS 790
QY 1307 PKRTYDMGGRVGRALSSASIEGLM-----GRAIPERHSPHILKEQHHI--- 1351
Db 791 POKTEVTASRLISAVQSPKVDPIMSDVTPTSPKKTATAPKDTSATLSLKSVPVAVTSL 850
QY 1352 -----RGSITOCIPR----- 1361
Db 851 SPPKAPVAPSNEATVPTPTSLKNALAAATPKETLATISIPKVTSPSPQKTPKSVSLKG 910
QY 1362 -----SYVEAQED-----YLREAKLLKRE-----GTPPP--pp 1388
Db 911 APAMTSKKATEIAAAGKDVSPSPKPEVPLLOHVPTSPKPSVSDTLGALTSPPPKGGP 970
QY 1389 PSRDLTEAVKTOALGPKLK-----PAHEGLVATVKEAGRSIETPR-----EELRHTPEL 1439
Db 971 ATLAETPTPKSPKPAASKKTPATPSPEGVTAVPLEIPPCSKKAPKTAAPKRESSATSS 1030
QY 1440 PLAPRLKBSGSIQ-----GTPLKYDTGASGTGSKKHVRSLSIGSPGRT----- 1483
Db 1031 KRAPKAVSKEIPSKGVTAVPLEISLPLKETSKSATPGEKS-----ASSPKRSPTAGP 1084
QY 1484 ---FPP-----VHPLDMADARALERACEESL-----KSRPGTA-----SSSGGSJAR--- 1524
Db 1085 KETPPGGVTAVPPEISLPPKETPQNATPNESLAASSQKRSKTPVPKETPPGGVTAMPLE 1144
QY 1525 ---GAPVIVPELCKPROSPLTYEDHAGFAPGHLPRGSPVTWREPTRLQSGSLSSKASQD 1582
Db 1145 IPSAQOKAPKTAVPKOIP--TPED-----AVTLGAGSLSPKK-----ASKTAAP 1187
QY 1583 RKLSTTPR-----ETAKSPHST-----VPEHHHPPIPSYEHLLRGVGV 1621
Db 1188 KEAPATPSGVTAVSGEISPSPKTSKTAAPKENSATLPPKSPKTAAPKETAPTSSEGV 1247
QY 1622 DLYRSHIPAFDPTIPRIGIPIDAAAAYLPHRLAPNTYPHLYPPYLIRGYPDTAALEN 1681
Db 1248 TAVPSEISPS--PPTASKGVV--TLTPKCAPNALAESPAASP-----KKVPKTAAPPE 1297
QY 1682 ROTIINDYTSQMHNTATAMAQADMRLGLSPRESSLALNYAAGPGRIIDLSQVPHLP 1741
Db 1298 TST-----TPSPQKIPKVG--PKEAS----- 1317
QY 1742 VLVPPPTPGTPATMDRLAYLPT-----APOPESSRRHSSSLSPGPGTHLTKPTTT 1791
Db 1318 -ATPSKTKPTAVKETSAPSEGVTAVPLEIPPSPRKAPKTAAP-----KETPA 1366
QY 1792 SSSERDRDRDRDREREKSILTSTTVEHAPITWRPGTEQSSGSSGGGSSSRAP 1851
Db 1367 PSPEG-----ATTAPVQIPPSPRKSKKA-----GSKETP 1396
QY 1852 ASHS-----HAHQSPISPRTODALQORPSVLHNTGMKGIITAVPSPKTVLIRSTSSP 1906
Db 1397 TTPSPGEGTAAPLEIPISKSKTSKASPKETLVTPSSKKLSQTVGPKETSLEGATAVPLE 1456
QY 1907 VPAATFEPAT-----HCLPGGLTLDGYVPTLMEPVLLPKEAPRVARPERPRADTGHAFIAK 1962
Db 1457 IPPSHKKAPKTVDPKQVPTSPKDAPTTLAESPSPKKAPKTAAPPSERVTT-----V 1510
QY 1963 PPARSGLEPASPSSKSGSEPRPLVPVSGHATITARTPAKNLAPHASPDPPAPPA--SASD 2020
Db 1511 PP-----EKPATPQKASGTASKVPV-----PAETQEVAVSSRETPTVPVPPVKN 1556
QY 2021 PHREKTSQKPSIQBELRLSLGYHGSSYSPEGVEVPSVSPSLTHDKLPLKHELDKS 2080

Db 1557 PSSHKTKSTIKELKE-----APATLPP--SPTKSPKIPSSKKAPR----- 1594
QY 2081 HLEGELRPKQPGVPLGGEEAHLPHLRPLPES--OPSSSPLLOTAPGVKGHQRVVTLAQHI 2139
Db 1595 ---TSAPK-----EFPASPSIKPVTTSLAQTAPPSLQKAPS----- 1627
QY 2140 SEVITQDYTRHHPOOLSAPLPAPLYS--FPGA-----SCP 2172
Db 1628 -----TTIPKENLAAPAVLPVSKSPAAPARASASLSPATAAPOTAPKEATTIPSK 1679
QY 2173 VLDLRRPPSDLYLPDPDHGAPARGSPHS-----EGGKS-----PEPNK 2211
Db 1680 KAAATETPIETATAPLEGAPKETSETSVKVLMSPPPKASSSRKASTLATPTLPSLKE 1739
QY 2212 TSVL-----GGEGDIEPVP--PEGTEPGHRSVAVPLLYRDGEQTEPSRMGSKSPG 2263
Db 1740 ASVLSPATSSGKDSHISPVSDACSTGTTTPQASEKL-----PSKKG----- 1781
QY 2264 NTSQPPAFPSKIT--ESNSAMVSKKQEKINKLTHNRNEPEYNIQSQTEIFNMPAITG 2321
Db 1782 ----PTAFTMLAAPAPESALAITAPIQKSPGANSNSASSKPCPDPSKKDKTGLPS--- 1834
QY 2322 TGLMITYRSQAVQEHASTNMGLEAIRKALMGKYDQWEESSPPLSANAFNLNA----- 2373
Db 1835 ---AVALAPOTVPEKDTSKAIETLLVSPAKGSDCLHSPKGVGSQVATPLAFTSKVPP 1892
QY 2374 ---SASLPAAMPITAADGRSDHTLT---SPGGGKAKVSGRPSRRKAKSPACPLASGDRP 2427
Db 1893 EAVSASV-APKAPAA-----SLTLAPSP-----VAPLPKQPLLESAPG--SVLES 1936
QY 2428 PSVSSVHSGEDCNRRTPPLNVRWEDRPPSAGSTPPFPYPLNLMRLQAGVNASPPPGGLPA 2486
Db 1937 PSKLPVPAEED--ELPPLI-----PPEAVSGGEPFQPILVN-----MPAPKPACTPA 1981
RESULT 8
A1819
proline-rich peptides 637K precursor, prostatic - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-May-1992 #sequence revision 13-May-1992 #text_change 05-Nov-1999
C:Accession: A42663; A41819; A31966; B20593; A20593
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
J. Biol. Chem. 267, 9884-9894, 1992
A:Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are
le exon.
A:Reference number: A42663; MUID:92250652
A:Accession: A42663
A:Molecule type: DNA; mRNA
A:Residues: 1-3706, 'I', 3708-4077, 'F', 4079-4155, 'S', 4157-5761 <DE2>
A:Cross-references: GB:M86514
A:Experimental source: ventral prostate
A:Note: sequence inconsistent with the nucleotide translation
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
submitted to the Protein Sequence Database, April 1992
A:Reference number: A41819
A:Accession: A41819
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-5762 <DE1>
R:Hemschoote, K.; Peeters, B.; Dirckx, L.; Claessens, F.; De Clercq, N.; Heyns, W.; W
J. Biol. Chem. 263, 19159-19165, 1988
A:Title: A single 12.5-kilobase androgen-regulated mRNA encoding multiple proline-rich
A:Reference number: A31966; MUID:89066721
A:Accession: A31966
A:Molecule type: mRNA
A:Residues: 3372-3540 <HEM>
A:Cross-references: GB:M20721; GB:J04188; NID:g206397; PIDN:AAA41950.1; PID:g554494
R:Peeters, B.; Heyns, W.; Bossyns, D.; Rombauts, W.
J. Biol. Chem. 258, 14206-14211, 1983
A:Title: Proline-rich polypeptides bound to rat prostatic binding protein. The primer
A:Reference number: A94675; MUID:84061859
A:Accession: B20593

A:Molecule type: protein
A:Residues: 2020-2057 <PEP>
A>Note: this peptide, designated proline-rich polypeptide V, can be found at several loci
A:Accession: A20593
A:Molecule type: protein
A:Residues: 2822-2859 <PE2>
A>Note: this peptide, designated proline-rich polypeptide IV, can be found at several loci
C:Genetics:
A:Introns: *status absent
A>Note: single copy gene with no introns
C:Superfamily: rat prostatic proline-rich peptides 637K precursor
C:Keywords: prostate; tandem repeat
F:1-26/Domain: signal sequence *status predicted <SIG>
F:27-5761/Product: prostatic proline-rich peptides 637K proprotein *status predicted <MA>

Query Match 3.9%; Score 512; DB 2; Length 5762;
Best Local Similarity 18.9%; Pred. No. 2.4e-10;
Matches 609; Conservative 342; Mismatches 1164; Indels 1112; Gaps 155;

QY 42 YQHSRDYASHLSPGSIQQRRLRSLLSEFQPGNRSQELHLRPES--HSYLPGLGKSE 99
DB 1311 YQEVSDQAEYTTSTV-----SFQP---LDQELTITSEATREPHHTVPQQT 1355

QY 100 MEFIKSRPRLELLPDLRLRSP-----LLATGQAGSDEL--TKDRSLTGKLE 146
DB 1356 IYVHTKHP-LVIHSEQTHPNPTVTVQPLDLTLNTPQTAEGELPOTLQDSTQIIE 1414

QY 147 PVS-----PPSPHTDPELEL-----V 163
DB 1415 PPTVVGVPIYEEVTVQTSQDAEYPPSPVSFQSLDGLTITPEPTEHITQKTVF 1474

QY 164 PPRLSKEELIQN-----MDRVDREITWVEQQLSKLKKQOQLEEAARPEPE 211
DB 1475 PMYTDVTLQVQSVQHLKPTGEGVQPLDLTLITPQTPPE--GELSTQVSTTONKEPH 1533

QY 212 KP-VSPPTLESKRSRSLVQIYDENRKKAAHRILEGLP-QVELPLYNQPSDTRIYHEN 269
DB 1534 KEVAPVPV-----YQAVTVPTPSQYAEYQ--KSQPLDLTLTVTSEPT-KEAYHST 1583

QY 270 IKINQMRKLLIYKRRNHARKQKQKFCQRYD-----OLMEALEKKVE 314
DB 1584 ISKNLAINPQVVIHQHPNPA-----EATVQPLDLLELTISSSLOPTAEGELLYSMQETV 1638

QY 315 RIENPRRAKESKRYEYKOPPELRKQELQRMOSRGVQSGLSMAARSEHEVSE 374
DB 1639 QISEPPKQV--TPPEYQEVAVP-----APVQDAKYPL---SSVLSNLSLDQELTSS 1688

QY 375 IIDGLSEQENLEKQROLAVIPPM-----LYDADQQRIFKIFNMGLMADPMKYKDRQVMN 430
DB 1689 ELGGAHQLTTPD---ETWVLPKDRQGIY-PDHDHKLHLNLTETVNTQPFHLETVQHP 1744

QY 431 MNSQEKETFRKFMQ-----HPKNFGLIASFLEK 461
DB 1745 TIEERSQIQKTTQITEPGKKVVPVLAQSEVITPMPILKETAPTPHSMALQSLDEK 1804

QY 462 -----TVAECVLYY-----YLTKKKEN 478
DB 1805 LTIHSPGWTQOAHNLKESKHTTGKILLDYAEPNMELELKHGLFLTKTTEATESEN 1864

QY 479 YKSLVRRSYRRRGKQQQQQQQQQQQ--QQQQQQQMPRRSSQBEKDEKEKEAEKEEK 536
DB 1865 TNQMTKSLQVATLTQNKKSMLPALVESQDESPPPNMSLQPLDQ-ELTLSSQPHGVVP 1923

QY 537 EYENKEDLLKEKTDTSGEDNDEKAVASKRKTANSQGR--KGRITRSMAN- 588
DB 1924 HIPNTPKILYHAEPPTGPFVEPPELFLTKTKRPVQGTATQMAASPCKEMVSRAPENK 1983

QY 589 -----EANSEEAITPOOSAEASLMELNESSR--WTEEMEETAKKGLLHGRNWSAI 637
DB 1984 EAVLSGPDQEDQESPPNMSLQSLDQELTLSSQPHGWTPHPNPT-----HGKIYLYH 2036

QY 638 ARWGSKTVSQCKNFYFNKKRQNLDELILQHQHKLKMEKERNARKKKKAPAAAEAAFP 697

DB 2037 AEPPTGPFVEPDDLFLFKTKSKPVWILTR-----TDKSRKEMVSQSKPYEAVLP 2088

QY 698 PVVEDEMEASGVSGNEEMVEABALHASGNEVPRGBCSGPATVNNSSDTSIESIPSPHTE 757

DB 2089 -----VHGEGQESR-----SPPNMSLSQLEGE 2111

QY 758 AAKDTGQNG--PKPATILCA-----DGPPPGPPTPP-----RRTSRAPIETP----- 797

DB 2112 LTLSSQPHGWVHPNTHGKIYLYHAEPPTGPFVEPDDLFLRTTKSKPVQGTATRMVKS 2171

QY 798 -----PA-SEATG-APTPP-----PAPP-SPS- 816

DB 2172 PEMVSLDPENKEAVFPQAQEGKGSPPNMSLQSLDHLFMSQPHGWPHPKPTDPK 2231

QY 817 -----APPPVVPKKEE-----ETAAAPVVEGEQKPPAAEELAVDTGKAEPEVKSCT 867

DB 2232 IYLYHAEPPTGPFVEPDDLFLRTTKSKPVQGTTEMAKSPKEMVSQTPYKEAVLSGPG 2291

QY 868 EAAEGPA-----KGDAEAAEAT-----AEGALKAEEKGSGR----- 902

DB 2292 EDQDESPPPNTSLKSDQEVAMSSQPHSGVPHPKTPGKIYLYHSIEPPPGPFVAKPTDLI 2351

QY 903 --ATTAKSSGA-----POSDSDSATCSADEVDEAEGGDKNRLLSRPP 942

DB 2352 LVKTTTKSKPAEWTPRIDKLLKEMVPHSPYEAEVFAHAGEGQDES-GSPNMPLOPLD 2410

QY 943 SLLTPTGDPRA-----NASPQ-----KPLD-----LKOLKORAAAIPPIQVT 979

DB 2411 QELTLSSQPHGWVHPNTHGKIYLYHAEPPTGPFVEPDDLFLRTTKSKPVQGSPEIA 2470

QY 980 K-----VHEPPREDAA---PTKPAPPAPP---ONLOPESDAPQPGSS 1018

DB 2471 KSPKEMVSQTPYKEAVLSGPDQDESPPNMSLQSLDQVMTWSSQPHSGVPHPP-KT 2529

QY 1019 PRGK-----SRSPAP-----PADKEAFAAEAKLQDGPWCWTSGLPFPVPREVIK-----A 1065

DB 2530 P-GKIYLSHSIEPPPGPFVKPTDLILVKTTKSKPAE--WT-----PRRIDKLLKEM 2577

QY 1066 SHAPDPAFAFVAPGHPPLGLHDTARVLPRLPPTISNPPPLISSAKHPSVLEIQIGAI 1125

DB 2578 VHSPE-----YEEAVFP-----AHGEGQDESPPNMPLOP-----LDQELTUS 2617

QY 1126 SOGMSVQLHVP-----YSEHAKAPGVV-----TMGLLPLPM--DPKKL 1161

DB 2618 SQPHGWVHPNTHGKIYLYHAEPPTGPFVEPDDLFLRTPKSKPVQGTPTQMAKSPPEM 2677

QY 1162 APFSGVKQEQLSRQAGPPESLGVPTAOEASVLRGTALGVSVPGGSITKGPSTRVPSDS 1221

DB 2678 VSLSPKNKETVFPQAQKQDESPPNLSLQSLDQEIWMSSQPHGWIPH-----PPN- 2729

QY 1222 ATTYRGSITHG-----TPADVLYKTIITRIGEDSPSLDRDREDSLPKG 1266

DB 2730 -----THGKIYLYHAEPPTGPFVEPDDLFLRTTKSKPMQGSPPROIDSKPEMTQ- 2780

QY 1267 HVIYEGKGHVLVSYEGGMSVTQCSKEDGRSSSGPHETAAPKRTYDMMEGRVGRAISSAS 1326

DB 2781 -----SPYEESLLPAHAEQESRAPHFSLOP-----LDQELSLSS 2818

QY 1327 IEGLMGRAPPERHSPHHLKEQHHRIGSITQIGIPRSYEAQEDYLRRKAKLLKREG----- 1382

DB 2819 -----HPHGWIHPHNPDPKIYLYHAEPPTGPFVEPDDLFLTKTKSKPVQGTATK 2869

QY 1383 TPPPPPPSRDLTEAYKTOAL-GP-----LKLKPAHEGLVAVTKAGRSIH--E 1427

DB 2870 TDKSPEDRVSTPYKEAVLSGPDQDESPPNMSLQSLDQELAISQPHGWIPHSPN 2929

QY 1428 IPRELRUTPELPLAP-----RPLKEGSIQ-----GTPUKYDTG 1462

DB 2930 APDKIYLYHAEPPTGPFVEPDDLFLTKTKSKPL--QGTPTQMAKSPKEMVSQTPYKEAD 2988

QY 1463 ASTTGSKHDKVRLSLGSPGRTFFPVHPLDMADALERACYEESLKSRP-----GTASSS 1518

```
Db 2989 LSAPGENRDE-----SPSSPNMSLHPLD-----QELSSSQPHGWIPHPNT 3030
QY 1519 GGSII-----ARGAPVIVPEL-----GKPRQSLPTYEDHGAPAGHLPRGSPVTMRE 1564
Db 3031 HGKIYLHYAEPTGPPVEPDDLFLKTKSKPVQSGPSQIDK-----SPKEVFS 3079
QY 1565 PPRRQOEGSLSSKASQ-----DRKLTSTPREIAKSPHSTVVEHHPH-PTS 1609
Db 3080 QSPSEESVLPAAEQEESRAPPHMSLQLDQKLT-----LSSHPHGWIP-HHPNTPGK 3133
QY 1610 PYEHLRLRGVSG-----VDLY--RSHIPLAFDPTSI PRG-----IPLDA 1645
Db 3134 IYLHYAEPTGPPVEPDDLFLKTKSKPVQSGPROVDKSPKEMFTQSPYEESVLPAA 3193
QY 1646 AA--AYILPRHLAPND-----TYPH--LYPP-----YLIRGYDPTAALENRTII 1686
Db 3194 EQEESRAPPHMSLQPLDQDLTLSSHHPHGWIHTHPNTPDKIYLHYAEPTGPPVEPDDL 3253
QY 1687 NDYITSOQHHTATAMAQADMLRGLSPRESLALNYAAGPRGIIDLQVPHLPVLVPP 1746
Db 3254 FLRTTKSKPAQWTPQMAKSPHEMVSLSPENKETV--FPAQKGQESISPPHMSLQPLD 3311
QY 1747 TGTATAMDLRLAYLPTAPQPESS--RHSSSPLSPGCP-----TLTKKPTTSS 1793
Db 3312 QDLTPSSPH--GWIPHPNTHGKIYLHYAEP--PTGPPVEPDDLFLKTKSKPVQVSP 3367
QY 1794 SPERDRDRDREREKSIITSTTVEHAPIRWPGTEQSGSGSGSGSGSGSSSRPAS 1853
Db 3368 RQIDKSPKEVFTQSPYEESVL-----PAAEQGEE----- 3398
QY 1854 HSHAQHSISPRTOQ-ALQQRPS--VLNTHGMKGII-----TAVESKPTVLR 1900
Db 3399 -SRAPPHMSLQPLDQDLTLSSHHPHGWIHTHPNTHGKIYLHYAEPPTGPPVEPDDLFLK 3457
QY 1901 TS-----TSSP-----VRPA-----ATFP-----A 1916
Db 3458 TSKPVHSGPROIDKSKHEMFTQSPYEESVLPAAEQEESRAPPHMSLQPLDQDLTL 3517
QY 1917 THCPILQ-----GTLGVVPTLMPEVLLP-----KEAPRVARPERPRADTGH 1957
Db 3518 SH-PHGWIHPHNTPKIYLHYAEPPTGPPVEPDDLFLKTKSKPVQSGPROIDKSPKE 3576
QY 1958 AFLAPPARSGLEPASSPK-----GSEPRLPVPVGHATIAATPAKNLAPHHA-SPD 2010
Db 3577 MFTQSPYEESVLPAAEQEESRAPPHMSLQPLDQDLTLSSHHP-HGWIPHPNTPDKIY 3635
QY 2011 ----PPAPPASADP---HREKTSKPSIQEELRSILGYHGS-----SYSP 2050
Db 3636 LHYAEPPTGP-FVEPDDLFLKTKSKP-----VHSGPRQIDKSPKEVFTQSP 3682
QY 2051 EGVEPVSPVSSPLTHDKGLPKH--LEELDK-----SHLEGEL--RPKQPGPVKLGGEAA 2101
Db 3683 EYEESVLPAAEQEESRA-PHMSLQPLDQDLTLSSHHPHGWIHTHPNTPDKIYL----- 3736
QY 2102 HLPHLRPLPESOPSSPLIQTAPGVKGHORVYTLAQHISEVITQYTRHHPOOLSAPLPA 2161
Db 3737 HYAEPPTGPPVEPDDLFLKTKSKPVHSGPROIDKSPKEVFTQSP-----PEYESVLP 3791
QY 2162 PLYSPFGASC-----VLDLRPPSDLYLPPD 2189
Db 3792 QAEDQESRAPPHMSLQPLDQDLTLSSHHPHGWIHTHPNTHGKIYLHYAEPPTGPPVEP 3851
QY 2190 -----HGAPARGSP-----HSGGKRSPPNKTSV--LGGGED-----GIEPV- 2225
Db 3852 LFFLTATKSKPVQWTPQTQIDKSREEMVQSPENEADIPRHGDGQDELNRPPIISLQPLD 3911
QY 2226 -----SPEG-MTEPGHSRSAYVPLLYRD-----GBQTEFSRMGSKSPGNTSQPPAFFSKL 2275
Db 3912 QELTLSSQPHGVVPPSTPDKIY-LHYAEPPTGPPVEP-----PDLFFLRT 3957
QY 2276 TESN-----SANKVSKQEIKNKLNTHNRNEPEN---ISOPGTETINMPAITGTGLMT 2326
Db 3958 TSKSPVQGTPTQLAKSPKEMVFQ-----TPEYKEAVLSAPGQDET----- 4000
```

```
QY 2327 YRSQAVQEHASTNMGLEAIIRKALMGKYDQ-WEESP-----LSANAFNP----- 2370
Db 4001 -----SSPNMSLSQLDQELTNSSQPHGMLPHPNTHGKIYLHYADPTGPFVEPPD 4051
QY 2371 ---LNASASLPRAMPTAADGRSDHDLTSTPGGGGKAKVSGRPSRRKAKSPACLASGDRP 2427
Db 4052 LFFLTATKSKPVQGTVTQMD-----KSPKQGLL----- 4079
QY 2428 PSVSSVHSGDCNRRUPLNRYWE--DRPSSAGSTPF-----PYNPLIMRLQAGVNASPP 2480
Db 4080 ----PAHQEAQDESPPNMSLQPVDOELSLSSQPHGWIHTHPNTPCKIYLH---YAEPP 4132
QY 2481 -----PPGL-----PAGSGPLAGPHHAWDEEPKPLLCQYETLSDSE 2517
Db 4133 TGFVEVPPDLFLKTKSKPLAGTPTQMAKSPKEMFS---QTLHKE 4176
```

RESULT 9

S36152

MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu

C:Species: Homo sapiens (man)

C:Date: 06-Jun-1995 #sequence_revision 17-Nov-1995 #text_change 15-Sep-2000

C:Accession: S36152

R:Irish, F.J.M.; Bougueleret, L.; Prieur, S.; Caterina, D.; Primas, G.; Perrot, V.; Ju Nature Genet. 3, 137-145, 1993

A:Title: Dense Alu clustering and a potential new member of the NkappaB family withi

A:Reference number: S36152; MUID:93272029

A:Accession: S36152

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1872 <IRI>

A:Cross-references: EMBL:Z15025

A:Note: in the authors' translation residues 32-34 are shown after residue 4 and, con

A:Note: the authors translated the codon AAT for residue 1000 as His

A:Introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1;

C:Superfamily: collagen alpha 1(IV) chain

Query Match 3.8%; Score 508.5; DB 2; Length 1872;

Best Local Similarity 20.9%; Pred. No. 9.4e-11;

Matches 428; Conservative 173; Mismatches 695; Indels 749; Gaps 94;

QY 663 DEILQOHKLKMKERNARKKK-----APAAASEEAAFPVVEDEMEASGVSGNEE 715

Db 331 EVDVTEKLKFFDEEDCRSDSEGAEGHRDSQSASGEER--PP-----EADGKKGN-- 379

QY 716 EMVEEAEALHAGNEVPRGECSPATVNNSSDTE-----SIPSPTEAAKDTGQNGP 767

Db 380 -----SPNSEPTTKTAAETSRPTEPGPPAPKPPPLPPPHRGAGNMGPPGD 428

QY 768 KPATILGADGPPGPPPTTP-----RTSRAPTE----- 795

Db 429 YP-----DRGGPCKPPAPEDEDAWRQRKQSSSEISLAVERRRRREERMOEERRA 484

QY 796 -----PTPASEATCAPTPPPAPPSPSPAPPVVPVKEEKEEETAAAPVVEEGEE 842

Db 485 ACAEKLRLDEKFGAPDKRLKAEPAPAPSTPAPPVAPVKE-----LPAPPA----- 533

QY 843 QKPPAAEEELAVDTKAEPEVKSECTEEAEAGPAKGDAAEATAEGALKAEKKGSGR 902

Db 534 --PP-----PASAPTEKEPEEPAQAPPAGSTPTP-----GVAA 565

QY 903 ATTAKSSGAPQSDSSATCSADEVD-----EAGGDKNRLSPRPSLLTTGTGPRANA---- 955

Db 566 APTLVSGGSGTSTSSGFEASVPQLPSKEGPEEPVPPPTTPPVVKEPKGDI GP 625

QY 956 -----SPQKPLDLKOLKORAAAIPIQVTKVHEPPREDAAPTKPPAPPAPPPQ 1003

Db 626 TRQPPSQGLGVKPYOKSLPPRFQROQQEOLLKQOQOHHQOQGSAPPTPPVPPSQPV 685

QY 1004 NLPQESDAPQQGSSPRGKRSRPAPPAKFAAEAKQLPGDPCWTSGLPPFPVPPREVI 1063

Db 6264 KKEKVPKVPBEPKPKVPKPKVVKVIRKMEBPLPAKVTEKHMQITQEBKVLVAVTKEA 6323
QY 861 PVKSECTE-----AERGPAKGDAAEAATAEG-ALKAEEKGGSGR 902
Db 6324 PKARVPEPKARVPEKVLKLPKREEPKAKVTEFRKRVVKEEKVSTEAPKREPQPIK 6383
QY 903 ATT---AKSSGAPQSDSATCSADVEAEGDGNRLLSRPLSTPGDPRAN-----954
Db 6384 EYTIMEKERAYTLEEAASVQREEEYEEYDYKEFEYEP---TEEDYQYEEYERE 6440
QY 955 -----ASPQKPLDLKOLKORAA-----AIP-----PIQVTKVHEPP 985
Db 6441 YERYEEHEEYITEPEKPIPVKVPPEPVTPKAPPAKVLKKAAPKVPVPIPKKLKP 6500
QY 986 -----REDA--PTKD-----APP-- 997
Db 6501 PKVPPEPKVPFEKIHISITREKEQVTEPAKVPKPKRVVAEKKVPVPRKEVAPPVR 6560
QY 998 APPPPQNLQESDA-----HAPDPSAFSYAPGH--PLPLGLHDTA 1092
Db 6561 VPEVPKLEPEEVAPEEVEVVTHVEEYLVEEBEYIHEREEFITEEBVVPIPVKVPVEP- 6619
QY 1017 SSPRKSRSAP-PAKKEFAFAEAKLQDGP-----PCWT 1050
Db 6620 RKPVPPEKKVPVPPKKEAPPAKVPKPKKEKVPVLPKPKKEPPAKVPVPEPKKVPPE 6679
QY 1051 SGLPPPPVPPEVIKASP-----HAPDPSAFSYAPGH--PLPLGLHDTA 1092
Db 6680 EKVVPVPK--VEAPPAKVPVPPKVPKVPKVPKAPKVE-APPKVPVPPKLIPEE 6736
QY 1093 RVLPRPPTISNPPPLISSAKHPSVLERIGAISQMSYQLHVPYSEHA--RAPVGPVTM 1150
Db 6737 KKTVPVKVKEAPPKPKKREP-----VPPVVALPOEEVEEVLFEIEEIVPEE 6783
QY 1151 GLP-----LPMPKPLAPSGV--KQQLSPRGQAGPPESLGPVTAQAEASVLRGALGSV 1203
Db 6784 VLPEEEVLPPEEEVLPPEEEVLPPEEEVLPPEEEVLPPEEEVLPPEEEVLPPEEEV 6833
QY 1204 PGGSITKGI-PSRVPS-----DSAITYSGSI-----THGTPADVLYKGITRILIGDPS 1253
Db 6834 PEEEVLPVKVPVPAVPPEIKKKVTEKKVVPKKEEAPPAKV-----PEVPK 6882
QY 1254 RLDRGREDSLPKGHVYIEGKGHVLSYEGGMSVTCQSKEDGRSSGPPHETAAPKRTYDM 1313
Db 6883 KVEEK-----IILPKEEVLPE-----VTEPEEPISEIEEIPPEPSIEVEEV 6929
QY 1314 MGRVGRATISSASIEGLMGRATIPPERHSPHLKEQHHRIGSITQGTIPRSYVAQEDYLR 1373
Db 6930 APPRVPEVIKKAPEA--PTVPVKVPEA-----PAKVSKKIPEKVPVVPVQ--KK 6976
QY 1374 EAKLLKREGTPPPPPSRDLTEAYKTOALGPLKPAHEGLVA-----TVKE----- 1420
Db 6977 EAPPAKVPVPPKVPKVKVLVP--KKEAVPPAKGTVLEKVSVAFRQVWVKERLEB 7034
QY 1421 AGRSIHEIPREELRHTPELAPRLKESITQGTPLKYDTGASTGSKKHDRVSLIGSP 1480
Db 7035 VEAEEVEIPEEEFEVE-----EYFEGEF-----HEVEEFIKLE 7070
QY 1481 GTFPPVPHLD-----VMADARALERACTEESLSKSRGTASSGGSIARGAP--VIYPELG 1534
Db 7071 QHRVEEHRVKVHVVIEVEEVEVEVEK-P-KAPPKGEPISEKIIPPKKPTKVVPKRE 7129
QY 1535 KPRQSP-----LTEDHGAPFAGHLPGSPVYTMREPTPLQEGSLSSSKASQDRKLTST 1588
Db 7130 PPAKVPVPPKVLVEEKVRVP---BEPRVPPTKVPEVLPPEKVPVPPPAKKPEAP 7186
QY 1589 PREIAKSHSTVPEHHPHIPSYEHLRLRGVSDVLRSHPIAFOPSTISPRGIPLDAAA 1648
Db 7187 PPKVPEAKVPEKVPKVPVPPK-----KPEVP-----PTKYPE-VPKAAVPE 7228
QY 1649 YLPLPHLAPNPTYPHLYPPYLIRGYPDTAALENRQTIINDYITSOOMHNTATAMAQRAD 1708
Db 7229 KKVPEAIPKPEP---PPEVPE----- 7248

QY 1709 MLRGLSPRESSLA-----LNYAAGPRGIIDLQVPH-----LPVLVP----- 1745
Db 7249 -----EPESPAPPKPEVPVVRVDEVPEKVPPEKVPFAAPPKPEVTPVKVPEAPKEV 7303
QY 1746 -----PTPGTPTATAMDRLAYLPTAPQPFSSRHSSSPLSPGPGTHLTKPTTSSSRE 1797
Db 7304 VPEKKVPVPPKPEVPPPTKVPEVKVAVPEKVPKPEAIPKPEP-----PPEVEEPEEV 7359
QY 1798 RDRDRDRDREREKSLISTTTTVEHAPTWRPGTEOSSSSSSSSSSSSSSSSSSSSPASHSHA 1857
Db 7360 ALEEPAEYVEEPEPAAPPQVTPVPPKNPV-----PEKKAPAVVAKK 7400
QY 1858 HOHSPTSPRTDALQORPSVHLNHTGMKGIITAVEPSKP-----TVLRSTSTSS 1905
Db 7401 PELPPV-----KVPEVKVEVPEKVPVLPVPPKPAKPAKVPDEVPEKVPKVVAV 7451
QY 1906 PYRPAATFPATHCPILGGTLDGYVPTLMPEVLLPKPEAPRVARPRADTGHAFKAPPA 1965
Db 7452 PKKPEV--PPAK-----VPEVPPKPVL--EEKPAVPVPERAESPPPEVY--EPEE 7495
QY 1966 RSGLEPASPSK-----CSEPRPLVPPVSGHATARTAKNL-----APHHA 2007
Db 7496 ETAPEEETAPPEEKVPVVAEEEEVEYPP-----PAVPEEPKKIIPKVPVIRKKEAPPK 7551
QY 2008 SDP-----PAPPASADPHREKTOSKPFST-----QLELRSLG 2042
Db 7552 EPEEKVIEKPKLPPPPPPPPAPPKEDYKEIFOLKAIPKKKVPENQVPEKVELTPLK 7611
QY 2043 YHGS-----SYSGEVPEVPVPS-----SPSLTHDKGLPKHLELDKSHLEGELEPK 2089
Db 7612 VPGGEKKVKLLPERKPEKEEVVLKSVLRKRPPEEPEKPEKLEKVKKPAVP---EPP 7668
QY 2090 QGPKY-----LGGEAAHLPHRLPLPESSPSSPLQOTAPGVKGQRVVTLAQHISEVI 2143
Db 7669 PKPVEEVEVPTVTKERKIPEPTKVPEIKP-AIPLAPEPKPKPEAEVKT-----I 7719
QY 2144 TDQYTRHHQQSLAPLAPLAPFPFG-ASCPVLDLRRPPSDLYPPPDHGAARGSPHSEG 2202
Db 7720 KPPVPEPEPTIAAPVTPVVGKKAEEKAEAKP-----GPIKGP----- 7764
QY 2203 GKRSPEPKNTS-VLGGGEGDIEP----- 2224
Db 7765 -KKTSPIEAERKKLRPGSGEKKPDEAPFYQLKAVPLKFKVKEIKDIILTESEFVGSSA 7823
QY 2225 -----VSPPEGMT-----EPCHSRSA-----VYPLLYRDEQTEPESRMGS 2259
Db 7824 IFECLVSPSTAITMWMKGSNTRESKHFADGDKDKLHIIDVQLSDAGEYTCVLRIGN 7883
QY 2260 KSPGNTS-----OPPAFFSKLTESNSAMVSK-----KQEKINKKLNTHRNEPEYNISQPG 2310
Db 7884 KEKSTAKLVVEELPVRFVKTLEEEVTVVGQPLYLSCLELNKERDVVWRDKGIIVKEPG 7943
QY 2311 TEIFNMPAIGT 2321
Db 7944 RIV---PGVIG 7951
RESULT 12
S37671
MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu
C:Species: Homo sapiens (man)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 15-Sep-2000
C:Accession: S37671
R:Bougueleret, L.
submitted to the EMBL Data Library, August 1992
A:Reference number: S37671
A:Accession: S37671
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1870 <BOU>
A:Cross-references: EMBL:Z15025; NID:g29374; PID:g29375
C:Genetics:

A:Map position: 6p21.3
A:introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 65
C:Superfamily: collagen alpha 1(IV) chain

Query Match 3.8%; Score 503; DB 2; Length 1870;
Best Local Similarity 21.2%; Pred. No. 1.5e-10;
Matches 431; Conservative 175; Mismatches 700; Indels 728; Gaps 97;

```
QY 663 DELLOQHLKMEKERNARKKK-----APAAASEAAFPVVVDEDEMEASGVSGNEE 715
Db 330 EEDVTEKLFDEEDGRDSEGAEGHSDQSASGEER--PP-----EADGKKGN-- 378
QY 716 EMVEEAELHAGNEVPRGECGSPATVNNSDTE-----SIPSPTEAAKDTGQNGP 767
Db 379 -----SPNSEPTTKTAWAETSRPTEPCGPAPKPLPPHPHGPAGNMGPGD 427
QY 768 KPATLGADGPPGPPPTTP-----RTSRAPTE----- 795
Db 428 YP-----DRGGPPCKPPAPEDEDAWRQRKQSSEISLAVERARRRREERMQEERRA 483
QY 796 -----PTPASEATCAPTPPAPSPSPAPPVVPKKEEETAAAPPVEEGEE 842
Db 484 ACAEKLRLDENKFGADPKRLKAEPAPPAAPSTAPPAPVKE-----LPAPPA----- 532
QY 843 QKPAAAEELAVDTGKAEPEVKSECTEEAEAGPAKGDKEAAEAATAEGALKAEEKGGSGR 902
Db 533 --PP-----PASAPTEKEPEEPAQAPPAQSTPTP-----GVAA 564
QY 903 ATTAKSSGAPQSDSATCSADEVD-----EAGGDKNLLSPRLLTPTGDPRANA--- 955
Db 565 APTLVGGGSGTSTSGSFEASVPQLPSKEGPEPEVPPPTTPVPVKPEKGDGIGP 624
QY 956 -----SPQKPLDLKOLKORAAIPIQVTKVHEPPREDAAATKAPAPAPPQ 1003
Db 625 TRQPPSQGLGYPKYOKSLPPRFQROQOOLKQOQOHOQOQSGASAPTPVPSPPPQV 684
QY 1004 NLOQESDAPQPGSSPRGSRSPADPAKFAFAEAQKLPDPPCWTSGLPPVPVPREVI 1063
Db 685 TL-----GAVPAPKAPPPPKA-----LYPG-----ALGRPPMPMPNF- 718
QY 1064 KASHPADPSAFSAPPQHLPLGLUHDTRAPVLPRPTISNPPPLISSAKHPSVLEROIG 1123
Db 719 -----DPRMMNIPYYDP-----RLQGRPPLDIFYPPGV-----HPSGLVPRER 757
QY 1124 AISOQMSVOLHVPYSEHAKAPVPTMGLPLPMDPKLA---PESGVKQEO---LSPRG 1176
Db 758 SDSRLSSE---PDRHAPAMLR--ERGTP-PVDP-KLAWGVDVETATPAEPRPLTSPLR 810
QY 1177 QAGPPESLGVPPTAQBAASVLRGTALGSVFGGSITKGIPTSRVPSDSAITYRGSITHGTAD 1236
Db 811 QAAEDDDKGM-----RSETPPVPP---PPPYLASYPGFPENGTTPCP 848
QY 1237 VLYKGTITRI-IGEDSPSLDRGREDSLPKGHVIEYEGKGHVLSVEGMSVTCQSKEDGR 1295
Db 849 -----PISRFPLEEPGRPL-----PMPGG-----SDEVAK 874
QY 1296 SSSGPPHETAAKPYTDMMGRVGRVRAISSAETGLMGRAPIPERHSPHLLKEQHHRGSI 1355
Db 875 IOTPPKPEKPEETAQLTGPAEAGKLPASR-----SGAGPPPPPRESTERFWRGPRGSS 930
QY 1356 TQGIPTSRVSEAOEDYLRREAKLLKREGTTP-----PPPPSRDLTBAYTKQALGPLKL 1407
Db 931 RRGIPP-----EPPGAPRRRAGPIKKPPPT-----KVEELPP--- 963
QY 1408 KPAHEGLVATVKEAGRSIETPRELHTPELAPRLK-----EGSITQ 1453
Db 964 KPEOG-----DETPK-----PPKPDPLKITKGLGPKETPPNGNLSP 1002
QY 1454 GTPLYKDTGASTG-----SKKHVDRSLGICSGRTFPVPHPLDVMADARALE 1500
Db 1003 APRLRDYSIERVGTSCRGGRGVEFARGRGFRGTGGRGR-----AQANSV 1052
```

```
QY 1501 RACYEESLKRPGTASSSGGSIARG--APVIVPELGP-----ROSPLTYEDHGA 1548
Db 1053 TESFEEM-----GVEVGQGDQITLLPEAALFARHCARVQSMRKSP-----SGA 1097
QY 1549 PFAGHLPRGSPV-----TMREPTRLQEGSLSSKASQDKRLTSTPREIAKSPHSTVP 1601
Db 1098 GSGAQQAARMPRVLWLLQTRRLPHPRREHSPRS-----RSPTTRSP 1140
QY 1602 EHHPHPISPYEHLLRGVSDVLYRSHIPL-AFDPTSIDRGIPLDAAAAYYLP-RHLAEN- 1658
Db 1141 TLHR---APARFTCPGV-----ESSLPEGAISPPRRREAPPQVCPGMSPPAKSLAPKK 1192
QY 1659 -PTYPHLYPP-----YLIRGPDTAALENRQTIINDYITSQOMHNTATAMAQADML 1710
Db 1193 PPTGP--LPPSKEPLKEKLIPLGPLSPVARGSGNSGVNMGEDGERPRRRHGRGAQQOK- 1249
QY 1711 RGLSPRESSLAL---NYAAGPRGIIDLSQVPLHPLVLPVPTTGTGTPATA---MDRLAYLPTA 1764
Db 1250 ---PPFRRLKQERENAAARGSEG-----KPSLTLPASAPGPEALTTVTA 1292
QY 1765 PQPFSSRHSSPLSPGPGTHLTKPTTSS-----SERERDRDE----- 1803
Db 1293 PAPRAAAKSPDLSNQNSDAQNEEWEETASESDFTSERRGDKEAPPPVLLTPKAVGTPEG 1352
QY 1804 -----RDRDREREKSIITSTTTTVEHAP IWRPGTEQSSGSG--SSGGGG 1845
Db 1353 GGGGAVPGISAMSRGDLISORAKDLSKFSQRPQWERNRPPGGKAGSGSGSGGAG 1412
QY 1846 GSSSRPASHSHAHQSPISPTQD--ALQORPSVLHNTGMKGIITAVEPSKPTVLRSTST- 1903
Db 1413 GPGGRTGPRGDKRSMWSPKNSRSPPEERP-----PGLPLPPPPSSSVFLRQV 1464
QY 1904 -SSVVRPAATFPPTHACPLGCTLGVYPTLME-----PVLLKPEAPVARPERPRADTG 1956
Db 1465 HSNPA-----GIOQALQALSRSQSVTAPGGHPR-HKPGPPQAPQG 1504
QY 1957 HAFIAKAPAR-----SGLEPASPSKSGSEPRPLVPVPSGHATITARTPAKNLAPHASP 2009
Db 1505 PS--PRPTRYEPQVNSGL---SSDPHFEPGPMVRGVG-----TPRSAG---VSP 1550
QY 2010 DPPAPPASADPHREKTKOSKPFSTQLELEL---RSIGYHGSSYSPEGVFVSPVSPSLTH 2066
Db 1551 FPP-----KRERPRKPELLQESLPPPHSSGFLGS--KPEG---PGQAESRDTG 1597
QY 2067 DKGLPKHL-----EELDKSHLEGELRPKOPGVKLGEEAHLPLRLP-ESOPS 2115
Db 1598 TEALTPHILNRLHTATSRKSYRPSMSEPMWEPLESPFEDVAGTEMSQSDSGVDLSGDSQVS 1657
QY 2116 SSPLIQ-TAP--GVKGHORVVTLAQHISEVITQDTRHHPOQL---SAPLPA-PLYSFPG 2168
Db 1658 SGPCSQRSRSPDGLKG-----AAGPPKPRGGSSPLNAVPCGPPG 1698
QY 2169 ASCPVLDLRRPSPDLYLPDPDHGAPARGSPHSEGGKRSPEPNKTSVLGGGEDGTEPVSFP 2228
Db 1699 SEPP-----RRPP-----PHDGRKELP-----REQPLPPGPIGTERSORT 1736
QY 2229 EGMTEPGHRSRVAVPLLYRDGEQTEPSRMGSKSPGNTSOPPAFTSKLTESNAMYKSKQ 2288
Db 1737 DRATEPGPIR-----PSHRPGPPVQF----- 1757
QY 2289 EINKLNTNHNNEPEYNISQGTETFINNPAITGTGLMYTYSQAOVHASTNMGLEAIRK 2348
Db 1758 -----GTSRDKDS-----DLRLVVGDSL-----KAEKE-----LTASVTE 1786
QY 2349 ALMGKYDOWEESPILSANAFNPLNASLASLPAAMPTIADGRSDHTLTSPGGGGKAKVSGR 2408
Db 1787 AIPVSRD-WEILLPSAAASA-EPQSKNLDGCHVPEPSSSGQRLYPEVFGSAG----- 1837
QY 2409 PSSRKAKSPAPGLASGRPPSVSVSHSGDCGNRRRTPLTNRVWEDRPPSAGSTPF 2462
Db 1838 PSSSOISGGSHGLS-----ITSKQWRLRPGTTPSLHPY 1869
```

RESULT 13

T13049 - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13049
R:Treisman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.
submitted to the EMBL Data Library, March 1998
A:Reference number: 217592
A:Accession: T13049
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2715 <TRE>
A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AAC06254.1
C:Genetics:
A:Gene: eld
A:Cross-references: FlyBase:FBgn0003013
C:Function:
A:Description: could act as a transcription factor antagonistic to the Wg pathway
C:Keywords: DNA binding

Query Match 3.8%; Score 502.5; DB 2; Length 2715;
Best Local Similarity 20.9%; Pred.No. 2.3e-10;
Matches 450; Conservative 189; Mismatches 752; Indels 763; Gaps 111;

QY 670 KLMKERNARRKKKAPAA-----ASEEAFPPVVEDEEAEASGVSGNEEMVEEAEA 723
DB 4 KIKSPQTOOQGGAPAPATPPSAGAPGATPT----- 39

QY 724 LHASNEVPRGCGPATVNNSSDTEIPS-----PHTAAKDTQONGKPKPATLGA 775
DB 40 -----SGPPTPNNSNGSDPSIQOQNVAPHYGA-----PPPGS-G 77

QY 776 DGPPPGP-----PTPRTSR-----APIETPASEATGAP----- 806
DB 78 PGPPPGPAPAAVMHYHLHQOQQHPPPHMQOQHGGAPPPPGGAPEHAPYKEEY 137

QY 807 -TTPPAP-----PSPAPPVVPVVEEKEEETAAAPPVEE---GEE 842
DB 138 HLPHPHPAYGRYHADPNMDPYRGQLPGKPP-----QQQHPHQOQPPQPGPGSP 193

QY 843 QKPPAAELAVDTGAEEVVKSECTEEAEGP-----AKGDAAEATAEGALKAEKK 896
DB 194 NRPPQOQVPGQPGPTLNSLLQSNPPPPQHRYANTYDQOAAASAAAAAQOQ 253

QY 897 EGSGRATTAKSGAPQSDSATCSADEVDEAEGDKN-----RLLSR----- 941
DB 254 QAG-GPPPPGHPGPPPHQPS-----PYGQOQGWAPPPRPSQPLGPSQOYRT 301

QY 942 -PSLLTPTGD---PRAN-----ASPQKPLDLKOLKORAAI-----PPIQVTK 980
DB 302 PPTNTSRGQSPYPAHQNSGSYSFSSFOOQOQOQOQOQOQOQOQOQOQOQOQOQ 361

QY 981 VHEPPREDAAFTKAPPAP-----PPQNL-----QPESDAPQPGSSPRGK 1022
DB 362 --OPPOQNTPTTSQXSPYQRYPTPEPLPAGSNHRTAYSTHYPEPNRPPGSSPSFG 419

QY 1023 SRSAPPADKAEFAEAKLPGDPCWTSGLPFVPPREVITKASP-HAPDPSAF-SYAPP 1080
DB 420 SGHPLPPASPH-HVPLPQOQPPPHVHVSAGGP---PP-----SSSPGHAPSPPSPQSP 471

QY 1081 GHPLPL-----GLHD---TARVLPRLPPTISNPPPLISSAKHPSVLERIGATS 1126
DB 472 SPHELIGQNSNDSSGHAHSGMSGPGTGNPQVMRTPTSGSS-----GSR 522

QY 1127 QGMSVOLHVPYSEHAK-----APV---GPVTMGLPLMDPKKLAPFSGYKQBLSPRQGA 1178
DB 523 MSPAVAQNHPISTRPASNSQSSGGMQOQPPVGAGGPPPPHPPHGMF-GGPPQOQOQOQOQ 581

QY 1179 GPPESLGVPYTAQEAELRGALGVSPPGSIYKIGISTRVPSDSALTYKSGITHGTADVL 1238
DB 582 SNSASSANSPOOT-----PPAPPPNMQNMNMTATPPPPP---QCAAGGYPMPH 629

QY 1239 YKGTITRIIGEDSPRLDRGREDSLPKGHVIYEGKKGHVLSYEGGMSVTCQSKEDGRSS 1298
DB 630 MHGGYK--MGPGQSPGAGYPPQOQ-----QYPPGNYPPRPQPPGAYAT 674

QY 1299 GPHETAAPAKRYDMMEGVRGAISSASIEGLMGRAIPPE--RHSPHHL---KEQHHRG 1353
DB 675 GPP-----PPPTSQAGAGGANSMPGSAQAGGYPRGMPNHTGQYPPYQVWPSPQOQTVPG 729

QY 1354 SITQIPRSYVEAQEDYLRREAKLLKREGTPPPPPSRDLTEAYKTOALGPLKLKPAHGS 1413
DB 730 ----GAPGAMVGNH-----VOGKGTPPPPVVG-----GP-----PPPOG 760

QY 1414 LVATVKEAGRSIHEIPRELRHTPELAPRLKKGSIQTQGTPLKYDTGASTYGSKKHDV 1473
DB 761 ----SGSPRLNLYL-KQLQHKGGYGSPTP-----PQG-PQGYGNG--PTGM----- 800

QY 1474 RSLIGSPGRTFPVH-----PLDVMADARALERACYEESLKSRLGTSASSGG 1520
DB 801 ----HPGMPMPGPHMGPPHGGPTNMGPTSTPPQSQMLQ-----GGQPGQCGASGG 847

QY 1521 STARGAPVIVPELGKPROSPLTYEDHGAPFAGHLPRGSP-----VTMRP 1565
DB 848 ----PESGGEH---ISQDNGISSG--PTGAAGMHAVTSVVTGDPGTGSMDEV 892

QY 1566 TPRLOEGSLS-SSKASQDRKLTSTPREIAKSP-----HSTVPEHHHPHPI-----SPYHL 1614
DB 893 S---QOSTLSNASASGEDPQCTTPKSRKNDPYQSOLAPPSTSPHVVHMGPGGGEY 949

QY 1615 LRG-----VSGDLYRSHIPLAFDP--TSIPRIGILDAAAAYLPRHLAPNTPHYLY 1665
DB 950 DMSPPNPNRPAGSPQVFNHVPQEPFRSTITTKKDSLCKLY---EMDDNPDR--- 1002

QY 1666 PLYLRGYPD--TALENROTIIINDYITSQOMHNTA--TAAQORADMLR-----G 1712
DB 1003 ----RGWLDKLRAMEERKRTPI-----TACTISKQPLDLYRLYYKERG 1045

QY 1713 LSPRESSALNAAAGPRGI-----IDLSQVPHLPVLVPPPT 1748
DB 1046 FVEVTKSTKWKDIAGLIGIGASSAAAYTLRKHYNLLTFECHFDRGDDIDPLIQOVEA 1105

QY 1749 GTPATAMDRLAYLTAPOQFSSRHSPLSPGPGPHLTKPTTTSSESRERDRDRDRDR 1808
DB 1106 GS-----KKTKAASVP--SPGG-GHLDAGTNTSG----- 1134

QY 1809 EREKSILSTTIVEHAPIWRCPTQESSGSSGSSGSSSPASHSHAHQHSIPRTQ 1868
DB 1135 ----SSNSQDSFPA--PGSAPNAAIDGYPGYGGSPYPGAS-----GPQPDYAT 1178

QY 1869 DALOORPSVLHN--TGKMGIIITAVEPSKPTVLRSTSTSSPVR-PAATFPPTHCPGLGTL 1925
DB 1179 AQOMQRPSONNPQTPHGAASAAVAGD-----NISVSNPFEDPTAAG-----GGPG 1225

QY 1926 DGVYPTLMEVLLPKAPRVARPERPRADTGHAFIAKPPARSGLEFPASSPSKSEPRPLV 1985
DB 1226 SGTGPG-----PQGP-----GPGAASGA-----GAVGAVGGGPQHP 1259

QY 1986 PVSCHAITATPA---KNLAPHASPPDPPAPASADPHREKTQSKPSIQEILERSLG 2042
DB 1260 PPHSPHTAAQOAOHQOHOHQHPLGPPPPFOOQOQOQOQOQOQOQOQOQOQOQOQ 1310

QY 2043 YHGSYSPE---GYEVPSPVSSSLTHDKLPLKLELDKSHLEGELRPKQPGVKLG 2098
DB 1311 -GGPPPAQOHCQGVPPSP-----QOHPVPAAGAYPPPG 1345

QY 2099 EAAHLPLRLPESQSSSPLLOTAGVKGHQVVTVAQHISEVITQDTRHHPOOLSAP 2158
DB 1346 SG-----YPTVSVNTPGSP-YPSQPGAY-----QYSSDQYNATGP 1381

QY 2159 LPAPLYSPFGACPVLDLRPPSDLYL-PP--PDHGAPARGS-----PHSEGGKRS 2206
DB 1382 PQQPGQPG-----QYPPQNRMVPPYGEAPEPTGANQYGPYSRYSQPPPG 1433

Db 1253 ----YEEILOTSIARMQASRDLAFTDKKKEKOFNAESAYMDPMKQNGGLTPGT 1308
QY 1665 YPYLIR--GYP-DTAALENKROITINDYITSQOMHNTATAMAQADMLRGLSPRESSLA 1721
Db 1309 SPTQLAAPVSPFTSTSSDGGNRVIDVRVQ-----1340
QY 1722 LNTAAGPRGLIDLSQVPHLPVLVPPT-----POTPATMDRLAYLPTAPQPFSSRHS 1773
Db 1341 -HPAKEPQEPKLRHSPASPSLASKEVGMTFSQCPGTPAT-----TAMAPCPASLP 1390
QY 1774 SSPLSPGGTHLTKPTTSSSERDRDRDRDREREKSSILSTTVEHAPWRGTEQ 1833
Db 1391 RGYMTAGPER--SPSTSTIHSYGPPPTANYGSOTEE-----LPHSPGSPAGSR 1440
QY 1834 SSGSSSSSGGG-----GSSS--RPASHSHAHOHSP--ISPR-TODALQ 1873
Db 1441 ASREKPLSGDGEVGPQPSRGYSYFTGSSPPLSPSTPSPESFSPKLCPRATAEFSTQ 1500
QY 1874 RPSVLNTHGKGIITAVEPSKPTVLKSTSTSSVRPAATFPFPAATHCPLGGTLDGVYPTLM 1933
Db 1501 TPS-----LTPSS-DIPRSVGTSPMVAQGTQTP--HRPSTPRLVWQOSSOE 1544
QY 1934 EPVL--LPKEAPRVARPERPRADT-----GHAF-LAKPPARSGL--EPASSP 1975
Db 1545 APVMVITLASDASSOTRMVHASASTSPLCSPTDSQASHYSQTTTPPSASQMPSEPAGPP 1604
QY 1976 SKGSEPRPLVPVPSGHATI---ARTPAKNLA-----PHASPD-PAPPASADPHRE 2024
Db 1605 ---GFPRASAGVDGLALYGWALPAENISLCRISVPGTSRVEGPRPPGTAVVDLRT 1661
QY 2025 KTOQKPFES--QBELELRSLGYHSSYSPEGVEPVSPVSPSLTHDKGLPKHLELDSH-- 2081
Db 1662 AVKPTPIILTDGMDLTSIAVEARKYG---LALDVPVGRQSTAVQPLVINLAQOQTHT 1717
QY 2082 -----LEGELRQKQPGVKLGEAAHLPHRLRLPESQSSPL-----2119
Db 1718 FLATATVTSITMASSVLAQKQPVVYGD-----PFSRLDFGQSGSPVCLAQVRQVEQ 1772
QY 2120 -LOTAP---GVKGHORVVTLAQH-----ISEVITQDYTRHHPOOLSAPLPAPLY 2164
Db 1773 AVOTAYRGPRGRPREAKFARNLPQVTPPLARRDILITOMGTAQSVLSKPGVPPE--- 1829
QY 2165 SFPGASCPVLDLRPPSDLYLPPDPDHGAPARGSPHSEGGKRRSPENKTSVLGGGE----- 2219
Db 1830 --PGAE-----PHRATPAELRAHALPQTRKPH---TVVQMGEGAAGT 1867
QY 2220 -----DGIEPVS-----PPEGWTE---P 2234
Db 1868 VTTLLPEEPAGALDLTGMRPESRLACCDMAYKFPFGSSCTGTFHPAPSPDKSVTDALP 1927
QY 2235 GHSRAVYPLLVRDGBOTEP---SRMGSKSPGNTSQ-----PPAFPSKLTESN--SA 2281
Db 1928 GQSSGFFYS--PRDPEPPELFRAGQVVGPGPHEQRPYPOGLPRLYSMSDNTMLAA 1985
QY 2282 MYKSKQEIKNKLNTHRN-----EPEYNI 2306
Db 1986 GLNYHAORIGQLFGQGRDSAVDLSLKHYSILGFPADGRYLGQGLQYGSFTDLRHPDILL 2045
QY 2307 SQP-----GTEIFNMPAIT-----2320
Db 2046 SHPLPMRPYSVSNYSNDRHYGPRGDAVGQFQASLAQYSAATTAREISRCAALNSMDQYG 2105
QY 2321 -----GTGLMYRSQAVQCHA---S'NMGLEATIRKALMGK--YDOWEESPPLSANA 2367
Db 2106 GRHGGSGGPDILVPQPOQ---HGFLNAPQGL-ASLSRGLGNPTYPYPGQSPGNLAQ- 2159
QY 2368 FNPLNASAS-----LPAAMPITAAAGDSRDLTSP-----2397
Db 2160 YGPAASOGTAVRQLLPSTATVRAADGMIVSTINTPIATLPIITQPASVLRPMVRGMYR 2219
QY 2398 ---GGGKAKVGRSPSRKAKSP--APGLASG-----DRPPSVSVHSEGDGNNRT 2443
Db 2220 PYGSGG---VTAVPLTSLTRVPMIAPRVPLGPAAGLYRYPAPSRFPASTI-----P 2267

QY 2444 PLTNRVWEDRP-----SSAGSTPFP-----YNPLIMRLOAGVMASPPPG 2483
Db 2268 PABGVPYLGKPAAKASGAGGPPRPPELPAGGAREEPLSTTAPPAVKEAPVAQAPPPPG 2327
QY 2484 -LPA-----GSGPLAGP 2494
Db 2328 QKPAGDAAGSGSGVLGRP 2346
RESULT 15
A43359
microtubule-associated protein MAP1A - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 13-Aug-1999
C:Accession: A43359; S22108
R:Langkopf, A.; Hammarback, J.A.; Muller, R.; Vallee, R.B.; Garner, C.C.
J. Biol. Chem. 267, 16561-16566, 1992
A:Title: Microtubule-associated proteins 1A and LC2. Two proteins encoded in one mess
A:Reference number: A43359; MUID:92355629
A:Accession: A43359
A:Molecule type: mRNA
A:Residues: 1-2774 <LAN>
A:Cross-references: GB:M8196; NID:g205537; PIDN:AAB48069.1; PID:g205538
A:Note: sequence extracted from NCBI backbone (NCBIN:111039, NCBI:P:111040)
R:Cravchik, A.
submitted to the EMBL Data Library, June 1992
A:Reference number: S22108
A:Accession: S22108
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 73-364, 'NRLRS', 370, 'QKN', 374, 'PSPKGL', 381-751, 'RSMMSOMNAORR', 764, 'D', 766, 'WLRNMCQPRQSP', 851, 'V', 853, 'NSL', 855, 'LPHRWLRN', 865, 'W', 867, 'HSQLPDGG', 877, 'Q', 87
A:Cross-references: EMBL:X66840
A:Experimental source: strain Sprague Dawley
C:Superfamily: microtubule-associated protein MAP1B
C:Keywords: microtubule binding; phosphoprotein
Query Match 3.7%; Score 492.5; DB 2; Length 2774;
Best Local Similarity 20.5%; Pred. No. 5.3e-10;
Matches 553; Conservative 278; Mismatches 1002; Indels 861; Gaps 132;
QY 77 ERSOELHLPESHYSYLPGLKSEMEFIESKRPRLELLPDLRPSPLLATQGPAGSEDLT 136
Db 398 DRAGKXHLK-EKISKLEE--KDKKEKKEKKELKKE-----EGRKEK 440
QY 137 KORSUTGKLEPVSPSPPHPTDPELELVPPRLSK---EELIQNMDRVDRITWVEQOISK 193
Db 441 KDAKKDEKRDKTPEVKKLSKPLDKPFTPEVRKTLKAKAPGRVKYDK-----GRA 491
QY 194 KKKQOOLEEAAKPPPEKVPSPPIESKHSLSV-----QIYDENRKAEEAAHRILE-- 246
Db 492 ARGEKLESEPTTPP-AQKGAAPPAVSGHRELALSSPEDLTQDFEELKREERGLLAQR 550
QY 247 --GLGQVQVELPLYNOPSD-TROYHENIKINOAMRKKLILYFKRRNHARKQWKQKFCORYD 303
Db 551 DTGLG---EKPL---PADATEQGHPSAAITQVTPSPGVLE-----584
QY 304 OLMEALEKKEVERIENNP-----RRRAKESKVEYEEKQFPEIRKQRELQERMQSRVQGR 358
Db 585 --GEHVEREKEVVPDPSGDKGTNRGPDPSGAEEVEKEKETWEERKQREAE-----631
QY 359 SGLSMSAAR--SEHEVSEIIDLSEQENLEKQMRQLAVIPPMYLDADQOIRKFINNGLM 416
Db 632 LGPENTAARESEAEYKEDVIEKAELEEMEE-----THPSD-----667
QY 417 ADPMKVYKDRQVNMNMWSEOEKETFEKFMQHPKPNFGLIASFLERKTVAECVLYYLTKN 476
Db 668 -----EGEETKAESEFYQKHTQOALKAPSKR-----694
QY 477 ENYKSLVRRSRRRGKSKOQQQQQ-----OQQQQQQQQQPMRPSOEEKDEK 525

Db 695 ---EALGGRDLGFQKAPKETAFLSSLATPAGATEHVSYIQDETIPGYSETEQTISDE 751
QY 526 EKEAEKEEB--KPEVENDKEDLLKKTDDTSGEDNDEKAVASKRGKTKANSQGR--K 579
Db 752 EIHDEDEPARPRPTSTYDLSGPEGCPFFASQAADSVAFPASSSKTYGAPETELTYPP 811
QY 580 GRITRSMANEANSEAITPQOASALAM-----ELNESSRTEEMETAKKGLLEHGRN 633
Db 812 NMVAAPLAEHEVSSATSITCDKLSSFATVAEDQSVAASLAPQTEETGKSSLL----- 866
QY 634 WSAIARMGSKTVSOCKNPFYKRONLDEILOQHKLMEKERNARRKKKAPAAASEE 693
Db 867 LDTVTSIPSRT-----EATQGLDYVPSAGTI-----SPTSSLEE 901
QY 694 ---AAPPVVVEDEMBASGVSGNEEEMVEEAEALHASGNVPRGECSPATVNNSSDTE 750
Db 902 DKGFKSP-----CEDFSVTESEK-----KGETVGRG--LSGKAVGKEEYV- 943
QY 751 IPSPHTEAKDTGONGPKPATLGADGP--PFGPPTPPRTSRAPTEPTPASEATGAPT 808
Db 944 ----VTSEKLSQYA----AVFAGPGHTLPPGEPALGVEERC--LSPDDSTVKMASP-P 992
QY 809 PRAPSPSAPP-----PVVPKEKEE-----ETAAAPV--EKEEQK--PAAAEALVD 854
Db 993 PSGPSSAHTPFHQSVEDKSPROFQEDSWGETHSPGVSKEDSEEQTVKPGPEGTSE 1052
QY 855 TGKAEPPVKSECTEAEAGPAKDAEAAEATAE--GALKAEKKEGGG-----R 902
Db 1053 EKGK-PTTSPQAQDMPVSIAGGTCTQLLPEQDKAIVFETGANGSLGAGTLPGEVR 1111
QY 903 ATTAKSSGAPQSDSATSABDEVDAEGDKNRLSPRSLTTPGDPANASPKQPLD 962
Db 1112 TSTEEATEPKQDEVLFTQDLSPEDAESLSVSVSPDT-----KQEATPRSPCS 1163
QY 963 LKOLORAAAIPQIVTKVHEPREDAATKAPAPPQNNLOQESDAPQOGSSPRCK 1022
Db 1164 LKEQPHKOLNWPVEDTQSLSFSESPSKET-SLDISKQLSPSLGTQFGEINLJK 1222
QY 1023 -SRSPAPPADKEAFAEAQKLPDPCMTSGLPFVPPREVAKAS-----PHAPDP--SAPS 1076
Db 1223 EERGPVMKAE-----DDSCHLAPVSIPEPHRATVSPSTDETPAGTLPGGSFS 1269
QY 1077 YA-----PGHPLPLGLH--DTARPVLPRPTISNP-----PP 1107
Db 1270 HSALSVDRKHSPGEITPGGHEFTSDSSLTKSPESLSSPAMEDLAVEWEGKAPKEKEPE 1329
QY 1108 LISSAKHP--SVLEQIGAISQMSV-----QLHVPSSEKAP----- 1144
Db 1330 LKSETRQKQKQILPEKVAVVEQDLIIHQKDGALDENKPGROODKTPEQKGRDLDEKDTA 1389
QY 1145 ----VGPVTMGLPLMDPKKLAPFGVKQEQLSPRGQ-AGPPESLGVPTAQEASVLRGTA 1199
Db 1390 AELDKGP-----EPKE-----KDLREDQOQAGPPAE-----KDKASBOROTD 1428
QY 1200 LGSVPGSGITKIPSTRVPSDSAITYRG-----ITHGTPADVLYKGTITRIIGED- 1250
Db 1429 LQOT-----QATEPRDRAQERRDSEKDKSLELRDRTPEE---KDRI--LVQEDR 1473
QY 1251 ----SPSLDR-----GREDSLPKGHVYE-----GKKGHVLSYEG--GM 1284
Db 1474 APEHSIPEPTQDRAPRDKGTDDKQKEEASEEKEQVLEQDKWALGKEGETLDQEARTAE 1533
QY 1285 SYTQSKED---GRSSSGPPHETAAPKRTY-----DMMEGRVGRAISSASTEGLMG 1332
Db 1534 OKDETLKEDTKOGKSSFFVEDTITTSKEIVLDQKSAEKADSVEQDQGALEKTRALGL-- 1591
QY 1333 RAIPPE-----RHSPLHLEKQHIRG-----SITQIGPRSYVEBAQEDYLRRKALLKREGT 1383
Db 1592 EESPAEGSKAREQEKYKWKQDVQGWRETSPTRGE---VGGQKE-----PVPAWEK 1642
QY 1384 PPPPP---PSRLT---EAYTQALGPLKLKPAHE---GLVAIVKAGRS---THEIPR 1430
Db 1643 SPEQEVRYWRDRDITLQODAYVRELSCDRKVWFPHELDQOGARPRYCEBERESTFLDEGPD 1702

QY 1431 EE---LRHTPELPLAP--RPLKEGSITQGTPLKYDVTGASTTGSKKHVDYRSLIGSPGRTF 1484
Db 1703 EGEITPLQHTPRSPWTSDFKQOEPLPKQGLEVERWLAEVGLPPEEDKLTSPFPEI 1762
QY 1485 -PPVHPLDVMADARALACRYEES-----LKSRRPGTAS-----S 1517
Db 1763 SPPASPPM--TGQVRPSAPGQESVPVDTSTAPMRNETTPSWLAEIPWPVPKDRPLPP 1820
QY 1518 SGGSTARGAPVIVPELKPQSPPLY-----EDHGAPFAGHLPRGSPV 1560
Db 1821 APLSPAPAPPTPAPPHPT--VPFSWGLAEYDSVVAQVEGAEELEGGYS---PLGKY 1875
QY 1561 TMREPTPRLOEGSSSKASQDKLTSTPREIA--KSPHSTVPERHPHPTISVYEHLLRGVS 1619
Db 1876 RKAEREREGGAGAPDSSSPKVPKEAGESLATRDTEQTEPEQREPTYPDE---RSFQ 1932
QY 1620 GVDLYRSHIPLAFDPTSIPRIGIPDAAAAYILPRHLA-----PNPTYPH 1663
Db 1933 YADIVEQMMLTGLGPACPTREPPLGASGDW--PPLSTKEEAAGCNTSAEKETSPASPO 1990
QY 1664 -----LYPPYLIRGYPDTA-----ALENQTIIINDYITSOQMH 1696
Db 1991 NLOSTPAFASYASLAGPAVPP---RQEPDGPVNVBSITPPAVPRAPISLSKDLSPPLN 2047
QY 1697 HNTATAMAQR-----ADMLRGL-----SPRE----- 1717
Db 2048 GSTVSCSPDRRTPSKETGCRGHWDGTDNDSDLEKAREOPEKETRSPSPHPHMPMGHSSL 2107
QY 1718 -----SSL-----ALNYAAGRGIIDISQVPHLPVLVPPPTGTATAM---- 1755
Db 2108 WPETEAYSSDSSHLGVSVRPSLDPPASAFGFSLLQAP--PQL--PSPAEPRSPACGSL 2163
QY 1756 ---DR-LAYLPTAQPTSSRHSS-----SPLSPGCP--THLTKPTT 1790
Db 2164 AFSGRDALALVPGTPT--RTRHDEYLVTKAPSLDSSLPQLPSPSPGGLLNLPRPAS 2221
QY 1791 TSSSERDRDRDRDREREKSILTTTTHAHPIWRPQTQSSGSGSGSGSGSSSR 1850
Db 2222 PALSE-----GSSSEATTPISSVAERFP--PGLE---AAEQSAEGLSGKE 2263
QY 1851 PASHSHAHQHSIPRPTQDALQORPSVLHNTGMKGIITAVEPFSKPTVLKSTSTSPVRPA 1910
Db 2264 SAAHS--LWDLTPLSPASASLDLAPA-----PA 2290
QY 1911 ATFFPATHCPLGGLDGVYPTLME-----PVLLPK-----EAPR-----V 1945
Db 2291 PAPAPAPGLP--GDLGDGTLPCRPECTGELTKKPSPLSPSGDHEANGPGETSLNPPGFVT 2349
QY 1946 ARPERPRADTGHAF-----LAKPPARSGLPEPASSPKSGSEPRPLVPPVPSGHATI 1994
Db 2350 ATAEEKEAEAPHAWERGSWPEGAERSSRPDTLLSSEQPLRPGKSSGG---PPCSLSSEV 2405
QY 1995 ARTP---AKNLAHP--HASP---DPPAPPASADPHREKTKQSKPFSIOELEL-----RS 2040
Db 2406 EAGPOGCACTDRPHCGELSPSLNPLPPS-----TODSDLSTEARLAGKGRRR 2456
QY 2041 LGYHSSYSP---EGVEPVSVPSSLTHDKGLPKHLEELDLSHLEGLRLRQPG---P 2093
Db 2457 VGRPGATGCPMADETPTTSASDSSGSDSDVPPEETECPSITAEEALDSEDEGDFLP 2516
QY 2094 V-KLGG-----EAHLPHRLPSPESQPSSSLQTLQATPGVKHQVRVTLAQHISEVITQ 2145
Db 2517 VDKAGVSGTHHPRGHDPPTPTLPDPRSP-----RP 2550
QY 2146 DYTRHHPOOLSAPLPAPLYSPFGASCVPVLDLRPPSDLYLPPPDHGAPARGSPHSEG--- 2202
Db 2551 DVCMDAPBGLSS-----ESGRVERLREKR-----PGRRAPGRAPKAPSPARRL 2593
QY 2203 ---GRSPPEPKNTSVLGGEGDIEVPSPGEMT---EPGH---SRSAVYPLYLRDQOT 2252
Db 2594 DIRGKRSTPGKGVDRRTSRTVPRPRSTPSQVTSAEKDGHSFMSKGLV-----NGLKA 2647

QY 2253 EPSRMGSKSPGNTSQPPAFFSKLTESNSAMVKSKKOEINKKLNTH-----NRNEP 2302
Db 2648 GSTALGSK---GGSGPPVYVDLAYIPNHCSGKTADQDFFRRVRASYVYVSGNDP 2698

Search completed: September 8, 2001, 14:36:54
Job time: 16781 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 14:32:18 ; Search time 53.39 seconds
(without alignments)
1614.929 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQWTRATEPRYP.....WDEPKPLCSQYETLSDSE 2517

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	13215	100.0	2517	1 NCR2_HUMAN	Q9Y618 h nuclear r
2	10987.5	83.1	2472	1 NCR2_MOUSE	Q9W442 mus musculus
3	4187	31.7	2453	1 NCR1_MOUSE	Q60974 mus musculus
4	4147.5	31.4	2440	1 NCR1_HUMAN	O75376 homo sapien
5	714.5	5.4	533	1 NCR1_RAT	Q9WUB5 rattus norv
6	526	4.0	2142	1 BAT2_HUMAN	P48634 homo sapien
7	492.5	3.7	2774	1 MAPA_RAT	P34926 rattus norv
8	443.5	3.4	2805	1 MAPA_HUMAN	P78559 homo sapien
9	428.5	3.2	2715	1 TRX2_HUMAN	Q9UNM6 homo sapien
10	424.5	3.2	2464	1 MAPB_MOUSE	P14873 mus musculus
11	412.5	3.1	1226	1 YCS3_YEAST	P25357 saccharomyc
12	410	3.1	2468	1 MAPB_HUMAN	P46821 homo sapien
13	400.5	3.0	1464	1 CAL1_HUMAN	P02452 homo sapien
14	391.5	3.0	1411	1 TCOP_HUMAN	Q13428 homo sapien
15	391.5	3.0	1460	1 CAL1_CANFA	Q9XSJ7 canis faml
16	391	3.0	3924	1 ANK2_HUMAN	Q01484 homo sapien
17	390	3.0	2944	1 CAL7_HUMAN	Q02388 homo sapien
18	389.5	2.9	3866	1 HRX_MOUSE	P55200 mus musculus
19	389	2.9	1670	1 CA34_HUMAN	Q01955 homo sapien
20	380	2.9	3669	1 HRX_HUMAN	Q03164 homo sapien
21	379	2.9	1685	1 CA54_HUMAN	P29400 homo sapien
22	378	2.9	1183	1 DRPL_RAT	P54258 rattus norv
23	373	2.8	1763	1 CA24_ASCSU	P27393 ascaris suu
24	367	2.8	1185	1 DRPL_HUMAN	P54259 homo sapien
25	358	2.7	1618	1 NEST_HUMAN	P48681 homo sapien
26	355.5	2.7	2842	1 APC_RAT	P70478 rattus norv
27	355	2.7	2090	1 N214_HUMAN	P35658 homo sapien
28	353	2.7	1453	1 CAL1_CHICK	P02457 gallus gall
29	352	2.7	1838	1 CAL5_HUMAN	P20908 homo sapien
30	350.5	2.7	2843	1 APC_HUMAN	P25054 homo sapien
31	348.5	2.6	1464	1 CAL3_MOUSE	P08121 mus musculus
32	348	2.6	1669	1 CAL14_MOUSE	P02463 mus musculus
33	348	2.6	3256	1 K167_HUMAN	P46013 homo sapien

RESULT 1

ID	NCR2_HUMAN	STANDARD;	PRT;	2517 AA.
AC	Q9Y618; Q9Y5U0; O13354; O00613; O15416;			
DT	01-OCT-2000 (Rel. 40, Created)			
DT	01-OCT-2000 (Rel. 40, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	NUCLEAR RECEPTOR CO-REPRESSOR 2 (N-COR2) (SILENCING MEDIATOR OF			
DE	RETINOIC ACID AND THYROID HORMONE RECEPTOR) (SMRT) (SMRTE) (THYROID-,			
DE	RETINOIC ACID-RECEPTOR-ASSOCIATED CO-REPRESSOR) (T3 RECEPTOR-			
DE	ASSOCIATING FACTOR) (TRAC) (CTG26).			
GN	NCOR2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM SMRT).			
RC	TISSUE=Pituitary;			
RX	MEDLINE=99178941; PubMed=10077563;			
RA	Ordentlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;			
RT	"Unique forms of human and mouse nuclear receptor corepressor SMRT.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM SMRT).			
RC	TISSUE=Cervix adenocarcinoma;			
RX	MEDLINE=99199215; PubMed=10097068;			
RA	Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;			
RT	"SMRTE, a silencing mediator for retinoid and thyroid hormone			
RT	receptors-extended isoform that is more related to the nuclear			
RT	receptor corepressor.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).			
RN	[3]			
RP	SEQUENCE OF 1023-2517 FROM N.A.			
RC	TISSUE=Cervix adenocarcinoma;			
RX	MEDLINE=96008552; PubMed=7566127;			
RA	Chen J.D., Evans R.M.;			
RT	"A transcriptional co-repressor that interacts with nuclear hormone			
RL	Nature 377:454-457(1995).			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM TRAC-1).			
RC	TISSUE=Fetal liver;			
RX	MEDLINE=96408715; PubMed=8813722;			
RA	Sande S., Privalsky M.L.;			
RT	"Identification of TRACs (T3 receptor-associating cofactors), a family			
RT	of cofactors that associate with, and modulate the activity of,			
RT	nuclear hormone receptors.";			
RL	Mol. Endocrinol. 10:813-825(1996).			
RN	[5]			
RP	SEQUENCE OF 428-613 FROM N.A.			
RC	TISSUE=Brain cortex;			
RX	MEDLINE=97369492; PubMed=9225980;			
RA	Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,			
RA	Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;			
RT	"CDNAs with long CAG trinucleotide repeats from human brain.";			

P34333 caenorhabdi
Q61315 mus musculu
P12036 homo sapien
P11087 mus musculu
P02461 homo sapien
P53420 homo sapien
P13942 homo sapien
P39060 homo sapien
O14497 homo sapien
P02458 homo sapien
P12270 homo sapien
P08640 saccharomyc

ALIGNMENTS


```
Db 961 PIGTPVSGYALYQRIKAMHESALLE-----EQROREQVDLECRSSTSPCST 1008
QY 1023 SRSPAPPADKEAFAAAQKLPDPCWTSGLPFPVPPPREVIKASHPADPSAFSAPPCH 1082
Db 1009 SKSP-----NRE-----W-----EVLQAPH-----QVITNLPEGV 1034
QY 1083 PLPLGLHDTARVLP RPPTISNPPLISAKHPSVLEIROGAISQ-----MSVOLHVPY 1137
Db 1035 RLP-----TTRTPRPPLIPSPSKTIVASEK-PSFI--MGGSISQGTPTGTYLSSHQAYP 1086
QY 1138 SEHAKAPGVPTMGLPLMDPKLAPFSGVKQEOISPRGQAGPPESLGVPTAQEAASVLRG 1197
Db 1087 QAPKPSVGSISLGLPQROESTKAAPLTIKOEESPRSONSQOPEGLLV-RAQHEGVVRG 1145
QY 1198 TALGSPVGSITKGLPSTRVPSDSAITVRGSIHTGTPA-----DVLKGTITRI-IG 1248
Db 1146 TA-GAVQESITRGTPASKISVETISLRLGSIQTQTPALPQAGITEALVKGPVSRMPLE 1204
QY 1249 EDSRLDRGREDSLPKGHVHIEGKGHVLSYEGGMSVTQCSKEDGRSSGPPHETAAPK 1308
Db 1205 ESSPEKV---REAAASKGHVIEGKSGHILSYDNKNA-----REGTRSPRTAHEMSL-K 1255
QY 1309 RYDMMEGRVGAIS-----SASTIEGLMGRAPPERHSPH-HLKEQHHIRGSIQTGIPR 1361
Db 1256 RSYEAVGSIKQGMMSRSPVSAPEGLITCALP--RGSPHSDLKERTVLGSGIMQGTPT 1313
QY 1362 SYVEAQEDYLRREAKLLKREGTPPPPPPSRDLTEAYKQALGPLKLKPAHEGLVATVKEA 1421
Db 1314 ATAESFEDGL-KYPOIKRES-----PIRAPEGAI-----TKGKP-YDG-IYTIEM 1358
QY 1422 GRSIHEIPREEL-----RHTPELPLAPRLKEGSIQTQTPKYDGTGASTGSKKHDRSL 1476
Db 1359 GRSIHEIPRODILTQESRKTPVQWOSTRPIEGSISQGTPIKFDNN-SQOSAIGHKXSL 1417
QY 1477 IGSPTGTPPVHPLDMAD-ARALERACE-----ESLKSRTGTAASSGGSTARGAPIV 1530
Db 1418 ITGPSKL--PRGMLIEVPIENIKVBERGYEDVKAGEPVRAHRTSVVSGSPVLRST---L 1472
QY 1531 PELGRPROSPLTYEDHGA-----PEAGHLPRGSPVTMREPTPLQEGSLSSSK-ASQDRK 1584
Db 1473 HEAPKAQLSPGLYDSSARRTPVSYQNTISRGSPMNR-----TSDVSSSKSASHERK 1525
QY 1585 LSTSTPRE-----IAKSPHSTVPEHHPHPTISPYEHLHVRGSGVDLYRSHIPLAPDPTSI 1640
Db 1526 SPLTPTQRESIPAKSPVPGVDPIVSH--SPEDPHRRSAAAGEVYRSHLPTHLDP-AMPFH 1582
QY 1641 IFLDAAAAYILPRHLAPNPTYPHLYPPVLYRGYPTDALEN-ROTIINDYLTSCQMHNT 1699
Db 1583 RALDPAAAYLQRLQSLPTGTPGTSQIOLY-----AMENTROTILNDYITSCQMQVNL 1633
QY 1700 ATAMAQRADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLPVPTPGTPATAMDRLA 1759
Db 1634 -----RPDVTGRGLSPREQGLPYPA-TRGIIDLTNMP--TILYPHAGGSTPPMPDRIT 1685
QY 1760 YLPTAPQPFSSR-HSSSPLSPGPHLHKPTTSSSERORDRDRDRDREREKESILYST 1818
Db 1686 YIPGQVTFPPRPYNAASLSPGHPTHL---AAAAGAERERERERERERERERERER 1742
QY 1819 TTVEHAP---IWRPCTGESSGSGSGGGGGSSRPASHAHQHSPISPRTQDA-LQOR 1874
Db 1743 ERIAAAPADLYLRPQSEQ-----PGRPGSHGVYRSPSP-SVRTQETILOQR 1787
QY 1875 PSVLHNTGMGIITAVEPSKPTVLRSTSTSPV-----RPAATFPPTATHCPLGGLTLDGVYP 1930
Db 1788 PSVFQGTNGTSVITPLDPTAQLRIMPLPSGGPSISQGLPASRYNTAADA-LAALYDAAS 1846
QY 1931 TLMPEVILPKAPR-----VARPERPADTGHAFAPARGGLEPASSPSK 1977
Db 1847 APQMDVSKTKESKHEARLEENLRSAVSEQQOQLEKNLEVERKRSVQCCTSSALPSG 1906
QY 1978 GSEPRPLV-----PPVSGHATITARTPAKN-LAPHHASDPDPAPPASADPHRE 2024
Db 1907 KAQHASVYVYSEAGKDKGPPPKRSRYEEELRTGKTTITANFIDVITITQIASDKDARER 1966
```

```
QY 2025 KTQSKPFSIOELELRSLGYHSGSYSPGVEPVSPVSSPSLTHDKGLPKHLELDKSHLEG 2084
Db 1967 GSQSDSS-----SSLSSHYETASDALEVISPASSPAPQEKPOAYQDPMVKANQEN 2020
QY 2085 ELRPQOPGVKLGGEAAHLPLHR-----PLPESQPSSSSPLLQT--APGVKGHQRVVTLAQ 2137
Db 2021 ESTROVEG-----LHHYRSQOESPSPQOQPLPSPSSQSEGMGQVPRTHRLITLAD 2071
QY 2138 HISEVITQDTRHH-POQLSAPL-----PAPLYSFCASCPCVLDLRRPPSDLYLPPDP--- 2189
Db 2072 HICQIITODFARNQVPSQASTTFTQSPSALSTP-----VRTKTSRYSSESQST 2123
QY 2190 --HGAPA-RGSPH-----SEGKRSPEPNKTSVLGGEDGIEPVSPPEGWTPEG-HSRSA 2240
Db 2124 VUHPRGPRVSPENLVDRSGRSPKSPERSHI---PSEYEPISPPQG---PAVHEROD 2177
QY 2241 VYPLYRDEQTEPRMGSKSPGNTSQPPAFESKLTESAMVSKSKQKOEINKKLTHNHN 2300
Db 2178 SMLLSQRGVDPAEQSRSDSRSPGISYLPSPFTKL-ESTSPMVSKKQEIFRKLNSGGG 2236
QY 2301 EPEYNISQGTIEFNWPAITGTLMTYRSQAOEHASTNGLEAIRKALMGKYDOWEE- 2359
Db 2237 DSDMAAOPGTEIFNLPAVTTSGAVSSRSHSPADPAS-NLGLIEDIIRKALMGSDDKVED 2295
QY 2360 -----SPPLSANAFNLNASASLPAAMPITAADGRSDHDTLTSPGGG-GKAKVSGRPSRRK 2413
Db 2296 HGVVMSPHV---GIMPGSASTSV-----VTSSEARDEGEPSPHAGVCKPKLINKSNRK 2347
QY 2414 AKSPAPGLA--SGDRPPSVSVHSEDCNRRTPLTNRVWEDRSPSSAGSTPPFPNPLMLR 2471
Db 2348 SKSPITGOSYLGTERPSSVSVHSEGDYHRQTP--GWAWEDRPSSTGSTQFPYNPLTIRM 2405
QY 2472 QGVNMASTPPPCGLPAGSGPL--AGPH---HAWDEEPKPLCSQYETLSDSE 2517
Db 2406 ----LSSTPTQIACSAITOAAPHQONRIWEREPAPLLSAQYETLSDSD 2452

RESULT 4
NCRL_HUMAN
ID NCRL_HUMAN STANDARD; PRT: 2440 AA.
AC 075376; Q90PV5; Q90Q18;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR).
GN NCOR1 OR KIA1047.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98393736; PubMed=9724795;
RA Wang J., Hoshino T., Redner R.L., Kajigaya S., Liu J.M.;
RT "ERO, fusion partner in t(8;21) acute myeloid leukemia, represses
transcription by interaction with the human N-COR/MSIN3/HDAC1
complex.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).
RN [2]
RP SEQUENCE OF 782-2440 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
RN [3]
RP SEQUENCE OF 974-2440 FROM N.A.
RX MEDLINE=99375328; PubMed=10444336;
```



```
Db 994 CGTSKSP-----W-----EVLQPAFH-----QLITNLP 1019
QY 1080 PGHPLPLGLDHTARPYLPRPTISNPPPLISSAKHPSVLERQIGALSQG-----MSVOLH 1134
Db 1020 EGVRLP-----TTRPRPPLIPSPKTTVASEK-PSFT--MGSSISQGTGTYLTSHNQ 1071
QY 1135 VPYSEHAKP-VGPVTMGLPLMDPKKLAFFSVKQEQLSPRQAGPPESLGVPTAQEAS 1193
Db 1072 ASYQETPKPSVGSISLGLPROESAKSATLPYIKOEFSRSPSONSQPEGLLV-RAHEG 1130
QY 1194 VLRGTAHSVPGSITKGPISRPVSDSAITYRGSTHGTTPA-----DVLYKGTIYR 1245
Db 1131 VVRGTA-GAIEGSIIRGTFTSKISVIESIPSLRGSITQGTPTALPGITPEALVGSISR 1189
QY 1246 ITGEDSPSLDRGREDLSLPGHVIYEGKGHVLSYEGGMSVTCQSKEDGRSSGPPHETA 1305
Db 1190 MPEDSSP--EKGREAAKGHVIEGKSHLSYDNKNA-----REGTRSPRTAHEIS 1242
QY 1306 AKRTYDMMEGRVRAIS-----SASIEGLMRAIPPERHSPH-HLKEQHHRIGSITQG 1358
Db 1243 L-KRSYESVEGNIKQCMSRSPSVAPLEGLICRALP--RGSPHSLDKERTVLSGSIQMG 1299
QY 1359 IPRSYEAQEDYLRRAKLLKREGTPPPPPPSRDLTEAYKTQALGPLKPAHEGLVAIV 1418
Db 1300 TPRATTESPEDGL-KYPKOIKRES-----PPIRAFEGAI-----TKGP-YDG-ITTI 1344
QY 1419 KEAGRSIHPIPREEL-----RHTPELPLAPRLKEGSIQTGPKLYDTGASTTGSKKHDV 1473
Db 1345 KEMGRSIHELPRODIIITOESRKTPEVQSTRPIEGSISQGTPIKFDNN-SGSAIKHNV 1403
QY 1474 RSLIGSPGRTFPVPHPLDYMAD-ARALERACYE-----ESLSRPGTASSSGSIAAGAP 1527
Db 1404 KSLITGPSKLSRGMPLEIVPENIKVVERGYEDVKAGETVRSRHTSVVSSGSPVLRS- 1462
QY 1528 VIVPELGRQSPLTYYEDGA-----PFAGHLPRGSPVMTWREPTPLRGSLSSSRASOD 1582
Db 1463 --LHEAPKAQLSGIYDDTSARTPVSYQNTMSRSGPMNRTSDVTIP-----PNKSTNHE 1516
QY 1583 RKLTSTPRE-----IAKPSHSTYPEHHPHIPISYVHELLRGVGVLYRSHIPFLAFDPTSIP 1638
Db 1517 RKSTLTPTQRESIPAKSPVGVDPVYSH--SPFDPHRGSTAGEVWHLPTQLDP-AMP 1573
QY 1639 RGILPD-AAAAIYPLPHLANPTYPHLYPPYLRGVPDPTAALEN-ROTIINDYITQQMH 1696
Db 1574 FHRALDPAAYLFRQLSFTPCYPSOQLY-----AMENTROTILNDYITQQMQ 1624
QY 1697 HNTATAMORADMRLCLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVPTTCTPATAMD 1756
Db 1625 VNL-----RPDVARGSLPREQPLGLUPA-TRGIIIDLTMNPP-TTLVPHPGGTSTPPMD 1676
QY 1757 RLAYLPTAPQPTSSR-HSSSPSLSPGPGTHLTKPTTSSSERDRDRDRDREREKSTIL 1815
Db 1677 RITYIPTQITPPRPNYSASMSGPHTHL-----AAAASAREREREREKERER-----TA 1730
QY 1816 TSTTVEHAPIWRPGTEQSGSGSGSGGGSSRRPASHAHQHSPISPRTQDA-LOQR 1874
Db 1731 AASSDL-----YLRPGSEQ-----PGRPGSHGYVRSPSP-SVRTQETMLQQR 1771
QY 1875 PSVLHNTGMKGIIITAVEPSKPTVLRSTSSPV-----RPAATFPATHPLGLTLDGVTP 1930
Db 1772 PSVFOGTNGTSVITPLDPTAQLRIMPLPAGGFSISGGLPASRYNTAADA-LAALVDAAS 1830
QY 1931 TLMPEVLLPK-----EAPRVARPERPRAD-----TGHAFLAK 1962
Db 1831 APQMDVSKTESKHEARLEENLRSSAAVSEQOQLEKTEVEKKSVOCLYTSSAPSG 1890
QY 1963 PPARSGLEPASS---PSKSEPRPLVPVPSGHATTARTPAKN--LAPHASDPDPAPPAS 2017
Db 1891 KP-----QPHSSVYVEACKDKGP--PPKRYEEELRTGKTTIIPAANFIDVLIIRQIAS 1943
QY 2018 ASDPREKTQSPFSTQIELELSLGVHSGSYPGEVPSVSSPSLTHDKGLPKHLEEL 2077
Db 1944 DKDAERGSQSSDSS-----SSLSSHYETPDAIEVISPASSPAPPQEKLTQYQPEVV 1997
QY 2078 DKSHLEGLRKPQPGPVKLGGEAAHPLHRLPLPESOPSSPLLQTAG-----VKGH 2129
Db 1998 KANOANDTRQTEGP-----LHHYRP-----QOESFSPQOQLPPSOAEGMGVPR 2045
QY 2130 QRVVITLAHISEVITODYTRHH-----POOLSAPLPAPLYSFPGA--SCPVLDLRRPPSD 2182
Db 2046 HRLITLADHICQIITDFARNQVSSQTPQO---PPTSTFQNSPALVSTPV---RTKTSN 2099
QY 2183 LYLPPPD-----HGAP-ARGSPH-----SEGKRSPEPNKTSVLGGEGDIEPVSPPEGM 2231
Db 2100 RYSPESQAQSVHHQRGSRVSPENLVDKSRGSRPGKSPERSHV---SSEYERISPPQ-- 2154
QY 2232 TEPGHSRSVAVPLLYRDGQTEBPMGSKSPGNTSOPPAFFSKLTESNMAVSKKQEIIN 2291
Db 2155 VPVVEHKQSDLLLSORGAEPAQRNDARSPGISVILPSFFTKL-ENTSPMVKSKQEIF 2213
QY 2292 KKLINHRNEPEYNIOPGTEIFNMPAITGTGLMTYRQAQVQEHASTNMGLEAIIRKALM 2351
Db 2214 RKLNSGGGSDMAAAQPGTEIFNLPVAVTSGSVSRGHSFADPAS-NLGLEDIIRKALM 2272
QY 2352 GKYDQHEE-----SPPLSANAENPLNASLPAAMPITAAAGRSDDHTLTSPGGGG--KA 2403
Db 2273 GSFDDKVEDHGVVMSQPMGV---VPGTANTSV-----VTSGETRREEGDPSHSGGVCKP 2324
QY 2404 KYSGRPSSRKAKSPAP--GLASGRDPPSVSVHSEGDCHNRRTPLTNRVWEDRPSASGTP 2461
Db 2325 KLIKSNRSRKSPIQCGYLGTERFSSVSVHSEGDYHRQTP--CWANEDRPSFGSTQ 2382
QY 2462 FYPNPLIMRLQAGVMASPPPPGLPAGSG-PLAGPH---HAWDEPKPLCSQYETILSDSE 2517
Db 2383 FYPNPLTMRM--LSSTPTPTIACAPSAYVNAAPHQONRIWEREPAPLLSAQYETILSDS 2439
RESULT 5
NCRL_RAT
ID NCRL_RAT STANDARD; PRG: 533 AA.
AC Q9WUB5; O70463;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR) (FRAGMENT).
GN NCOR1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99371771; PubMed=10441327;
RA Boutell J.M., Thomas P., Neal J.W., Weston V.J., Duce J., Harper P.S.,
RA Jones A.L.;
RT "Aberrant interactions of transcriptional repressor proteins with the
RT Huntington's disease gene product, huntingtin.";
RL Hum. Mol. Genet. 8:1647-1655(1999).
RN [2]
RP SEQUENCE OF 476-528 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=99421707; PubMed=10491148;
RA Schuler M.J., Buehler S., Pette D.;
RT "Effects of contractile activity and hypothyroidism on nuclear hormone
RT receptor mRNA isoforms in rat skeletal muscle.";
RL Eur. J. Biochem. 264:1982-1988(1999).
CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE
CC ABSENCE OF LIGAND (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
```



```
Db 1573 KPEGPGQAEASDRTGTEALTPHWNRLHTATSRKSYRPTSMPEMPLSPFEDVAGTEMS 1632
Qy 2038 LRSLYGHSSYSPGVEPVSPVSSSLTHDKGL-----PKH-----LEELDASHLEG 2084
Db 1633 QSDSGVDLSGDSQVSGPCSRSSP-----DGLKGAAGPPKPRPGGSSPLNAVPCGPPG 1688
Qy 2085 ELRPKQPGVPLGGGAHLPLRPLPESQPSPLLTQAPGVKGHORVVTTLAQHISEVIT 2144
Db 1689 SEPPRRPPAPHDGRKELPREQPLP-----PGDIGTER-----SQR-----T 1726
Qy 2145 QDYTRHHPQQLSAPLAPLYSPGASCVPDLRR-----PPS-DL 2183
Db 1727 DRGTEPGPIRPS-HRPGPVQF-GTSDKSDSLRVVGDLSKAEKELTASVTEAIPVSRDW 1784
Qy 2184 YLPDPDHGAPARGSPHS---EGGKRSPPDN-----KTSVLGGGEDGI- 2222
Db 1785 ELLP---SAAASAEPOSKNLDGHCVPFESSGQRLYPEVFGSAGPSSQISGSGHGLS 1841
Qy 2223 -----EPVSPPEGTEPGHRSRAV-VPLLYRDGEQTEPSRMGSKS 2261
Db 1842 ITSQWRLRPPTSLHPRYSQLYLPFGPAPPSSALLSGVALKGQFLDFSTWQATELGKLP 1901
Qy 2262 PONTSQPPAFF-----SKITESNAWVSKKQOEINKLNTNHRNPEYNIQPGTEI 2313
Db 1902 AGGVLYPPPSFLYSPAFPCSPPLDTSLLQVRO-----DLSP-SDF 1941
Qy 2314 FNMPAITG--TGLMYRQAOVQEHASTNGLLEAIIRKALMGKYDQWESPPLSANAFNPL 2371
Db 1942 YSTLPQGGQSGFLSPGAPAOQ-----MLLPWVDSQLPVVNFGLSPAPP---PAPPPL 1992
Qy 2372 NASASLPAAMPITAADGRSDHPLTSPGGGKAKVSGRSPSRKAKSPAPGLASGDRPPSVS 2431
Db 1993 SLLPVGPALQPPSLA-----VRPPPAATRVLPSPA-----RPPPAS 2029
Qy 2432 SVHSE-----GDCNRRPTLNRVWEDRPSAGSTPPFPYNPLI--M 2469
Db 2030 LGRALHPVELKPFQDYQKLSNLGPGSSRTPPTGTSFSGLSRLKATPTSTYSGVFRTQ 2089
Qy 2470 RLQAGVMASP-----PPGLPAGSGPLAGPHHAWDEEP 2502
Db 2090 RVDLYQQASPPDALRWIPKWPWERTGPPREGPSRR-AEEP 2128

RESULT 7
MAPA_RAT
ID MAPA_RAT STANDARD; PRT; 2774 AA.
AC P34926;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2].
GN MAP1.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92353629; PubMed=1379599;
RA Langkopf A., Hammatback J.A., Mueller R., Vallee R.B., Garner C.C.;
RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
RT one messenger RNA."
RL J. Biol. Chem. 267:16561-16566(1992).
CC -!- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN THE FILAMENTOUS
CC -!- CROSS-BRIDGING BETWEEN MICROTUBULES AND OTHER SKELETAL ELEMENTS.
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
CC APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
CC THEIR MORPHOLOGY.
```

```
CC -!- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
CC FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC -!- PWM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC -!- PWM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B.
CC -!- SIMILARITY: TO MAP1B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M83196; AAB48069.1; -.
CC PIR; A43359; A43359.
CC CHAIN ?2465 2774 MAP1 LIGHT CHAIN LC2.
CC Microtubules; Repeat; Phosphorylation.
CC DOMAIN 309 496
CC DOMAIN 336 541
CC REPEAT 336 338 1.
CC REPEAT 415 417 2.
CC REPEAT 420 422 3.
CC REPEAT 424 426 4.
CC REPEAT 427 429 5.
CC REPEAT 431 433 6.
CC REPEAT 436 438 7.
CC REPEAT 440 442 8.
CC REPEAT 444 446 9.
CC REPEAT 449 451 10.
CC REPEAT 539 541 11.
CC SEQUENCE 2774 AA; 299526 MW; 3DEF74427BA9D7D7 CRC64;
CC
CC Query Match 3.7%; Score 492.5; DB 1; Length 2774;
CC Best Local Similarity 20.5%; Pred. No. 2.4e-09;
CC Matches 553; Conservative 278; Mismatches 1002; Indels 861; Gaps 132;
Qy 77 ERSQELHLPESHVYLPGLGKSEMEFIESKPRLELLPDLPLRSPPLATQOPAGSEDLT 136
Db 398 DKAGKKHLK-EKISKLEE--KKDKKKKKELKKELKKKE-----EGRKKEK 440
Qy 137 KDRSLTGKLEPVSPSPHTDPELELVPLRSK---BELIQNMDVRDREITMVEQISKL 193
Db 441 KDAKDKARKDTKPEVKLSKPDLPKPTPEVRKTLKAKAPGRVKYDK-----GRA 491
Qy 194 KKQOQLEEEAAKPEPEKPVSPPIESKHSLSV-----QIIYDENRKKAAAHRILE-- 246
Db 492 ARGEKELSESPRTTP-AQKGAAPPAVSGHRELALSSPEDLTQDFEELKREERGLLAQR 550
Qy 247 --GLGPQVELPLYNQPSD-TRQYHENIKINQAMRKKLLILYFKRRNHARKQKQFCQYD 303
Db 551 DTGLG---EKPL---PADATEQGHPSAAIQVTPSGPVLE-----584
Qy 304 QLMALEKKVRIENNP-----RRRRESKVREYKQFPEIRKQELQERQMSRVGQRG 358
Db 585 --GEHVEREKEVVDSPGDKSTNRGPDGSGAEVEKEKETWEERQREAE-----631
Qy 359 SGLSMSAAR--SEHEVSEIIDLSEQENLEKOMRQLAVIPPMPLYDADQQRKFINMNGLM 416
Db 632 LGPNTAARESEAEVKEDVIEKALEEMEE-----THPSD-----667
Qy 417 ADPMKVYKDVORVMNWSQEKETREKFMQHPKFNGLIASFLERKTVACVLYVYLTKN 476
Db 668 -----EEGETKAESFYQKHTQALKASPKSR-----694
Qy 477 ENYKSLVRRSRRRGKSOQOQOQO-----QOQOQOQOQPMRPSQOEKDEK 525
Db 695 ---FALGGRDLFGOGKAPEKETASFSLSLATPAGATEHVSIQDETIPGYSETQITSE 751
Qy 526 EKEAKEEEE--KPEVENDKEDLLKEKTDGSDGNDKEKAVSKGRKRTANSQGR--K 579
```



```

RESULT 8
MAPA_HUMAN
ID MAPA_HUMAN STANDARD; PRT; 2805 AA.
AC P78559; Q15882; Q12973;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2].
GN MAP1A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97001161; PubMed=8812494;
RA Fink J.K., Jones S.M., Esposito C., Wilkowski J.;
RT "Human microtubule-associated protein 1a (MAP1A) gene: genomic
RT organization, cDNA sequence, and developmental- and tissue-specific
RT expression."
RL Genomics 35:577-585(1996).
RN [2]
RP SEQUENCE OF 134-419 FROM N.A.
RC TISSUE=Fetal muscle;
RA Channilkulchai N., Pasturaud P., Richard I., Auffray C.,
RA Beckmann J.S.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1607-1883 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95356255; PubMed=7629894;
RA Fukuyama R., Rapoport S.I.;
RT "Brain-specific expression of human microtubule-associated protein 1A
RT (MAP1A) gene and its assignment to human chromosome 15."
RL J. Neurosci. Res. 40:820-825(1995).
CC -!- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN THE FILAMENTOUS
CC CROSS-BRIDGING BETWEEN MICROTUBULES AND OTHER SKELETAL ELEMENTS.
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
CC FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC -!- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B.
CC -!- SIMILARITY: TO MAP1B.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
EMBL; U38291; AAB41132.1; -
DR EMBL; U38292; AAB41133.1; -
DR EMBL; 247038; CAB87104.1; -
DR EMBL; U14577; AAB1362.1; -
DR MIM; 600178; -
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ?2490 2805 MAP1 LIGHT CHAIN LC2.
FT DOMAIN 309 496 LYS-RICH (BASIC).
FT DOMAIN 415 541 9 X 3 AA REPEATS OF K-K-[DE].
FT REPEAT 415 417 1.
FT REPEAT 420 422 2.
FT REPEAT 427 429 3.
FT REPEAT 431 433 4.
FT REPEAT 436 438 5.
FT REPEAT 440 442 6.

```

```

FT REPEAT 444 446 7.
FT REPEAT 449 451 8.
FT REPEAT 539 541 9.
FT CONFLICT 134 135 VV -> IP (IN REF. 2).
FT CONFLICT 249 249 A -> G (IN REF. 2).
FT CONFLICT 263 263 V -> A (IN REF. 2).
FT CONFLICT 311 311 S -> G (IN REF. 2).
FT CONFLICT 335 336 ST -> AK (IN REF. 2).
FT CONFLICT 353 353 S -> A (IN REF. 2).
FT CONFLICT 357 357 S -> A (IN REF. 2).
FT CONFLICT 364 364 Q -> K (IN REF. 2).
FT CONFLICT 414 419 EKDKKE -> KKRNS (IN REF. 2).
FT CONFLICT 1650 1650 C -> W (IN REF. 3).
FT CONFLICT 1690 1690 S -> A (IN REF. 3).
FT CONFLICT 1714 1714 V -> G (IN REF. 3).
FT CONFLICT 1869 1869 E -> A (IN REF. 3).
FT CONFLICT 1879 1883 GTPEY -> AHSRV (IN REF. 3).
SQ SEQUENCE 2805 AA; 306051 MW; 9027AF693EFFFBE3A CRC64;

Query Match 3.4%; Score 443.5; DB 1; Length 2805;
Best Local Similarity 21.5%; Pred. No. 9.4e-08;
Matches 420; Conservative 188; Mismatches 756; Indels 587; Gaps 91;

QY 18 RYPHSL- - - - - YPQIARTHTDVGLEYOHHSRDYASHLSGSIQIP 61
DB 1139 RYPDRSLSPDEXESLSVSPSPDTXXQEPKSPCLXEQXLHDKRWX- - - - - VSP 1191
QY 62 QRRRPSLLSEQPNGNRSOEL- - - HLRPESHSLP- - - - - ELGKSEMEFTESKRPRLELLP 114
DB 1192 EDTQSLSEESPSKETSLDVSSKOLSPESLGTLOFGLNLGKEEMGHLMQAEIDSHHTA 1251
QY 115 DPLLRPSLLATGP- - - - - AGSEDLTKDRSLTKLEPVSPSP-PHTDPELELVPP 165
DB 1252 -PMSVPEHAATASPTDGTTFYSAQTITDD- - - - - SLDRKSPASSFSHTSPS- - - - - 1299
QY 166 RLSKEELIONMDRVREITVMVQOISKLLKQOQLLEEAAPKPEPEKVPSP- - - PIESKH 223
DB 1300 - - - - - GNGKYLPGAITSDEHI- - - - - LTXDSSFSKSPESLXGPAIXDIAIKW 1342
QY 224 RSLVQIYD- - - ENKKAAFAHRIEGLGPQVELPLYNQPSDTROYHENIKINQAMRKL 280
DB 1343 EDKVPGLKDRTSEKKEPEPEKDEVLQOKDKTLEHKEVVEPKDTAIYQK- - - - - DEALHVK- 1397
QY 281 ILYFRRRHARKQWKQKFCQR- - - YDQLMEALEKKVERIENNNRRRAKESKVREYEQF 337
DB 1398 - - - - - NEAVKQKALEKGRDLQXDTALEQDKALE- - - PKDKDLEKDKALEQDK 1448
QY 338 PEIRKQRELQERMOSRVGQSGLSMSAARSEHEVSEIIDLSEQENLEKQMRQLAVIPP 397
DB 1449 IPXEKDKALE- - - - - QKDTALEQDKALEPKDKDL- - - - - EQKDRVLEQKEKIPE 1493
QY 398 MLYDADQORIKFINNGLMADPMKYKDRQVNMWSEQEKETFREKFQHPKFNGLIASF 457
DB 1494 EKDKALDQKRVSE-HKAPEDTVAEMKDRDL- - - - - EOTDKAPEQKHQAOEQK- - - - - KV 1543
QY 458 LERKTVAECVLYLYITTKKNENYKSLVRRSYRRRKSQOQQOQQOQQOQQOQQPMPRSSQ 517
DB 1544 SEKKDOALEXYWALGQDKDALEXNI- - - - - QALEENHQTQBSLQVEDKTRKPKMLE 1597
QY 518 BEKDEKEKEKAEE- - - EKPEVENDKEDLKE- - - - - KTDTS 554
DB 1598 EKSPKVKAMEEKLEALLEKTALGESLVQEGRAEQEKEKYWRGQDVQBCQTSPT 1657
QY 555 GEDNDEKAVASKGRKTANSQGR- - - KGRITFSMANEENSEE- - - - - AI 596
DB 1658 EEPAGEQKELAPAWEDTSPQDNRYWRGREDVSLQDTYWRLELSCEKRWKWPFPHELDVOGA 1717
QY 597 TPQSAELASMELNESSRWTEEMETAKGLLEHGRNNNSATARMVGSKTVSCKNFYNY 656
DB 1718 RPHYTEERESTFLDEG- - - PDDEQEV- - - - - LREHATR- - - - - SPWASDFKDF- - - - - 1758
QY 657 KRONLDEILQOHLKMEKERNARKKKKAPAAASEEAAFPVPVEDE- - - - - EMEASGVSG 712

```


Db 1759 -----QESSPOKGLVER-----WLAESPVLGPEEEDKLTSPFEIISPPA 1800
QY 713 NEEEMVEAEALHAGSNEVPRCEGSGPATVNNSDTESIPSPHTEAAKDTGQNGKPPAT 772
Db 1801 SPPEWVGQ-RVPSAGQESP-----IPDKLPHMKNEPTTPSWLADIPWPVVKDRPLPPAP 1856
QY 773 LGADGPPGPTTPPTSRAPTEPTPASEATGAPPPP-----APPSAP 818
Db 1857 L---SPAGPPTP-----APESHTPAFWSWCTPEYDSVVAQVGAEEGPGYSPGLK 1907
QY 819 PVVPEKEEETAAPVEEGEBQPPAAELAVDTGKAE-EPVKSECTEAEAE-----872
Db 1908 DYRKVXGEREEGRAPDKSHXKVPKSHATTEPEQTEPEQREPTPYPDERSFOY 1967
QY 873 -----GPAKDAEAAEATAGALKAEKGGSGRATTAK-----SSGAPQSD 916
Db 1968 ADIYEQMMLTGLGPCACPTREPLPGAOWPPCLSTYKAAAGNRTSAEKLSPSPKSLQ 2027
QY 917 S-----SATCSADEVDEAEGGDKNRLSPRLTPTGDPRA---NASPOKPLD---L 963
Db 2028 SDTPTFSYAALAGPTVPPRKEGP-----SMEPSLTTPAVPPRAPILSKGSPPLNGNLL 2082
QY 964 KOLKORAAIP-----PIQTKVHEPPREDA--APTKPAPAPPPP--PQNLQEPES 1009
Db 2083 SCSPPDRSPSPKESGRSHWDDSTSDSELEKGAQEAQSPSPHPPTPMGSPTLWPET 2142
QY 1010 DAPQOPG-SSPRGKRSAPPADKFAFAEAKL-----PGDP--PCWTSGLFP- 1055
Db 2143 EAHVSPPLXHLUGXAR---PSLDFPASAGFSSLEXAPQXPSPAEPRAPC--GSLAFS 2197
QY 1056 -----PVPPR-----EVIKAS-----PHAPDPSAFSPAGPHPLPLGLHDTAR 1093
Db 2198 GDRALAPAGPTTRYDEYLEVTKASLSDSLPQLPSPSS-----PCXPLLSNLPRAS 2252
QY 1094 FVLPRPTISNPPPLISSAKH---PSV---LERQIGAIQSGMSVQLHVYSEHAKAPGVG 1148
Db 2253 PALSEGSSEATTPISSVAERFSPLEAAEQESGELDPGEPAHXLWDLTPLSPAPPA 2312
QY 1149 TMGL---PLPMDP-----KKLAPSGVYKQELSPRG--QAQPPESL 1184
Db 2313 SLDLALAPSLPDGMDGILPCHLECESEATEKSPFQ-VPSECAANGPTETSPNPPX 2371
QY 1185 GVPTA---QEASVLRGTALGSPVGSITKIPSTRVPSDSAITYRGSITHGTPADVLYKG 1241
Db 2372 PAPAKAENEAAAXPAWEGAWPEGAERSSRPTXLSPEQPV-----2413
QY 1242 TITRIIGEDSPRLDRGREDSLPKGHVYEGKKHVLVYEGGMSVTCQSKEDGRSSGPP 1301
Db 2414 -----CPXG-----GSGGPP 2423
QY 1302 HETAAPKRTYDMGRCVGRATISSASIE---GLMGRAIPPERHSPHLKEOHIRGSIQTG 1358
Db 2424 -----SSASPEVEAGPQGCXTEPRH-----RGELSPS 2451
QY 1359 -----IPRSYVEAQEDYLRREAKLKR-----EGTPPPPPPSRDLTEAYKTOALGPL 1405
Db 2452 FLNPLPSPIDD--RDLSTEEVRLVGRGRRRVGGTGXPXVPTDETPTPTSASDSG--2507
QY 1406 KLPKAEGLVATVKAGRSIHIEPREELRHPELPLAPRLKEGSITQCTPLKYDTGAST 1465
Db 2508 -----SSQSDSDVPE-----TECPSPITAEAAALDDEGDFLPVDKXGV 2548
QY 1466 TGSKK---HDVRSILGSPGRTFPVPHPLDVNADARAL-----ERACYEESLSRGT 1514
Db 2549 SETHHPRGHDPPLQPDQXRPSP- RPDVCMADPEGLSSESGRXERLXXKEKVQVRVG- 2606
QY 1515 ASSSGGSIARGAPVIVPELGRQSPFLYEDHGAFFAGHLPGRGSPVTWRETPRLQEGSL 1574
Db 2607 -----RRAP-----GDKX--PV-----SPXRRLXLRGK---RSPTPGKSDXR 2639
QY 1575 SSSKASQDRKLTST--TPREIAKSPHSTVPEHHPHPI--SPYEHLLRGVGVDLVYRSHIPL 1630

Db 2640 VSRXPXRSRSTXSVQTPAE-EKDGHSPMSKGLVNGLKAGXPALSKGSGGAPVY---VDL 2695
QY 1631 ADPTIS-IPRGPLD-----AAAYLPRHLAPNPYPHYLPYLLRGYPDTAAL-----1679
Db 2696 AYIPHWGSKTADLDFRRVRASYVVSNDP-----ANGEPSRAVLDALEGG 2743
QY 1680 -----ENRQ---TIINDYITS-----QQMH 1696
Db 2744 KAQWGENLQVKTLIPTHTDTEVTREWYQOTH 2774
RESULT 9
TRX2 HUMAN
ID TRX2 HUMAN STANDARD; PRT: 2715 AA.
AC Q9UMN6: Q9UK25; Q95836; Q9Y669; Q9Y668; O15022;
DT 01-OCT-2000 (Rel. 40; Created)
DT 01-OCT-2000 (Rel. 40; Last sequence update)
DT 01-OCT-2000 (Rel. 40; Last annotation update)
DE TRITHORAX HOMOLOG 2 (MIXED LINEAGE LEUKEMIA GENE HOMOLOG 2 PROTEIN).
GN TRX2 OR HRX2 OR MLL2 OR KIAA0340.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RA Angrand P.O., Valvatne H., Jeanmougin F., Adamson A.,
RA van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,
RA Lamerdin J., Chambon P., Lossos R., Stewart A., Aasland R.;
RT "Mammalian trithorax- and ASH1-like proteins: putative chromatin
RT regulators which contain PHD fingers and SET domains."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RA Lamerdin J.E., McCready P.M., Adamson A.W., Burkhardt-Schultz K.,
RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Ganes J., Danganan L.,
RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,
RA Carrano A.V.;
RT "Sequence analysis of a 1 Mb region in human 19q13.1";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 816-2715 FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
RN [4]
RP SEQUENCE OF 111-2715 FROM N.A. (LONG ISOFORM).
RC TISSUE=Testis, and Leukocyte;
RX MEDLINE=20105772; PubMed=10637508;
RA Huntsman D.G., Chin S.-F., Muleris M., Batley S.J., Collins V.P.,
RA Wiedemann L.M., Aparicio S., Caldas C.;
RT "MLL2, the second human homolog of the Drosophila trithorax gene, maps
RT to 19q13.1 and is amplified in solid tumor cell lines";
RL Oncogene 18:7975-7984(1999).
RN [5]
RP PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).
RC TISSUE=Placenta, and Bone marrow;
RX MEDLINE=99339983; PubMed=10409430;
RA Fitzgerald K.T., Diaz M.O.;
RT "MLL2: A new mammalian member of the trx/MLL family of genes.";
RL Genomics 59:187-192(1999).
CC -!- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC -!- TRUNCATED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS.
CC ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS,
CC SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL

Db	2102	RARP-PEDLP	-----SETV--DFVL-----KNLGGP-----	2122
Qy	2165	SFFGASCPVLDRRRPSDLYLPPDPHGAP	--ARGSPHSEGGKRS-PFNKTSVLGGGEDG	2221
Db	2125	GDGAG	-----PREESLP-----APPLANGSPSGGLTASPADPRTTFAWLPGAPG	2171
Qy	2222	IEPVS	-----PPEGMTPE--GHRSAYVPLYLRDGBQTEPSRMGSKSPGNTSQPPAFF	2272
Db	2172	VRYLSGLSGPAPEPPKPAKTSKILVNLKLGQVFKMAGEGEPVPP	-----VKQPPL--	2220
Qy	2273	SKLTESNAWVSKKQEKINLNTNHNREPEYNISQPCTEIFNPAITGTLMTYRSQAV	2332	
Db	2221	-----PPTISPTAPISWTLPQGLLGVLPVV-----	2246	
Qy	2333	QEHASTNMGLEAIRKA	-----LMGRKYDQWEEESPPLSA-----NAFNPLNASASLPA	2379
Db	2247	-----GVRPAPPPPPPLTLVLSGSPASPPQAIIRVKRVSTFSGRSPAPPY	2295	
Qy	2380	AMPIITAADR-SDHTLTSPGGGKAKVGRPSRKKAKSP-APGLASGDRPPSVSSVHSEG	2437	
Db	2296	KAPRLDEGEASEDTPQVPLG	-----SGGFSRVKMTPTTVRGVLDLDRGEPAGEESP	2350
Qy	2438	DCNRRTPLTNRVYEDRPSAGSTPPFNPPLMLRLQ	-----ACGVMS-----PPPCLPAGSG	2489
Db	2351	PLQERSPLI-PIPEDGPPQVDPGP	-----PDLLLESQWHHYSGASSEEEPPSPDKNQA	2406
Qy	2490	P-LAGPH 2495		
Db	2407	PKRTGPH 2413		
RESULT 10				
MAPE_MOUSE				
ID	MAPB_MOUSE	STANDARD;	PRT; 2464 AA.	
OC	PI4873;			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-APR-1990 (Rel. 14, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	MICROTUBULE-ASSOCIATED PROTEIN 1B (MAP1.2)			
DE	THE MICROTUBULE BINDING DOMAIN OF MICROTUBULE-ASSOCIATED PROTEIN 1B (MAP1.2)			
GN	MAP1B OR MTAP5.			
GN	MAP1B OR MTAP5.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
ON	NCBI_TaxID=10090;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SWISS WEBSTER; TISSUE=Brain;			
RK	MEDLINE=90094539; PubMed=2480963;			
CC	Noble M., Lewis S.A., Cowan N.J.;			
RT	"The microtubule binding domain of microtubule-associated protein 1B (MAP1B) contains a repeated sequence motif unrelated to that of MAP2 and tau".			
RT	J. Cell Biol. 109:3367-3376(1989).			
CC	-1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.			
CC	-1- PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN STABILIZING MICROTUBULES.			
CC	-1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE WITH MAP1A AND MAP1B PROTEINS.			
CC	-1- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE KKEE AND KKEIV, REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES BOTH IN VITRO AND IN VIVO.			
CC	-1- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION OF MAP1B.			
CC	-1- SIMILARITY: TO NEURAXIN.			
CC	This SWISS-PROT entry is copyright. It is produced through a collabora			

--
CC
CC OF MARAD:
CC
CC -I- SIMILARITY: TO NEURAXIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
CC

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X51396; CAA35761.1; --
 DR PIR; S07549; QRMSP1
 DR MGI; 97179; Mtap5.
 DR InterPro; IPR000102; --
 DR Pfam; PF00414; MAP1B_neuraxin; 10.
 DR PROSITE; PS00230; MAP1B_NEURAXIN; 7.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN ? 2464
 FT DOMAIN 589 787
 FT LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 KREE AND KKEI/V REPEATS).
 12 X 17 AA TANDEM REPEATS.
 FT DOMAIN 1865 2068
 FT REPEAT 1865 1881 1.
 FT REPEAT 1882 1898 2.
 FT REPEAT 1899 1915 3.
 FT REPEAT 1916 1932 4.
 FT REPEAT 1933 1949 5.
 FT REPEAT 1950 1966 6.
 FT REPEAT 1967 1983 7.
 FT REPEAT 1984 2000 8.
 FT REPEAT 2001 2017 9.
 FT REPEAT 2018 2034 10.
 FT REPEAT 2035 2051 11.
 FT REPEAT 2052 2068 12.
 SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDBDA87 CRC64;

Query Match 3.2%; Score 424.5; DB 1; Length 2464;

Best local Similarity 19.3%; Pred. No. 3.4e-07;

Matches 521; Conservative 314; Mismatches 996; Indels 869; Gaps 114;

QY 68 LLSFQPGNRSQELHLPESHSLPELG-----KSEMEFTESKRPRLELLPDPLL 118
 DB 174 LLSSTHPANKASLT-----FCPEGDWKNLDRHNLQDFINIKLNSAIIIP----- 221
 QY 119 RPSLLATQAPAGSBDLTKDRSLTKLEVPSPPHDTPLELVPP-----RLSKE-- 170
 DB 222 -----EMGLSEFTEYLS-----ESVEVPSPFDI-----LEPPTSGGFLKSKPCC 262
 QY 171 -----ELIQNMDRVDREITVVEQOI----- 190
 DB 263 YIFPGGRGDSALFVNGFNLINGSERKSCFWKLTIRHLRDV-----SILLTHIGDDNLPG 319
 QY 191 --SKLKKKQOOOLEEAAKPPPEPEKVPSPPIESKHSRLVQIIYDENRKKAAEAHRIEGL 248
 DB 320 INSLQRTAELEEE-----RSQGSTNSNDWKNLI 350
 QY 249 GPQVELPLYNQSDTRQYHENIKINQAMRKKLILFKRNHARKQWKOKFCORYDQLMEA 308
 DB 351 SPDLGVFLNVPENLKDPEPNIKMRKSIEEACFTL-----QYLNKLSMKPEPLFRS 401
 QY 309 LEKKEV-----RLENNPRRAKESKVREYEQFPEIRKOR--ELQERMOSRYGQ 356
 DB 402 VQNTTEPVLIFQKMGVGLKEMVLPNPKVSSKEMQFMQWQGTCTNKDKAELLILPNQOEVDI 461
 QY 357 RGSGL-SMSAARSEHVS-----EIIDGLSEQENLEKQMRQLAV----- 394
 DB 462 PISYLTSSVSLIWHPANPAEKIIIRVLPFGNSTQYNILEGLEKHLKHLDFLQPLATQKDL 521
 QY 395 -----IPMLYDADQORIKFINMGLMDPMKVKYKDRQVMNWMWSQEKE--TFREKFMQHP 448
 DB 522 TQGVTPPVQVVKLOR-----ADRSRESLPATPKPVASKSVRKESKEETPEVTNTKSQVETP 578
 QY 449 KNFGLIASFLERKTVACVLYLYLKKENYKSLVRRSVRRGKSKQOQOQO--QOQOQOQO 506
 DB 579 K-----VESKEV-----LYKKDKPVKTESKPSVTEKESKEEQSPVKAEEVAK 623

QY 507 QOOQPMRSPSOBEKDEKE--KEKEAEKEBEKPEVENDKEDLLKEKTDGTDGSDNDKEAVA 565
 DB 624 QATESKPKVTKDKVVKELKTKLEKKEKPK---KEVVKEDKTPKKDKPKRKEEVK 679
 QY 566 SKGRKTANSQGRR-----KGRITRSMANEANSEEAITPQOSAEALSMELN 610
 DB 680 KEIKKEIKKEERKELKKEVKETPLKDAKKEVKKEKKEKKEK--EPKKEIKKISKDIK 738
 QY 611 PSSRWTEEMETAKKGLLEHGRNWSAIARWGSKTVSQCKNFYFNKKRQNLDEILOQUK 670
 DB 739 KST-----FQSDTKP-----SALKPKVAKKEESTKK-----EPLAAGK 772
 QY 671 LKMEKERNARRKKK---APAAASEEAAFPVVEDEMEASGVSGNEBEMVEEAALHA 726
 DB 773 LKDKGKVKVIKKEGKTTAAATAVGTAAATAAATAAAGIAASG-----PVKELEA--- 822
 QY 727 SGNEVPRGECGPATVNNSSDTEIPSPHTEAAKDTGONGPKPPATLGADGPPGPPPTPP 786
 DB 823 -----ERSILMSPEDL--TKDFEELKAEIDVAKDI-----KPQLELIED----- 860
 QY 787 RRTSRAPTEPTPASEATCAPTPPPAPSPSPAPPPVPVVKKEEKEEETAAAPPVE-----BGE 841
 DB 861 ----EELKETQPGEAY-----VIQETEVSKGSASPDEGITTTEGE 899
 QY 842 EOKPPAAEELAVDTGKAEEPVKSECTEAEAGPKAGKDAEAAEATAEGALKAEEKEG--- 898
 DB 900 GECEOTPEEL-----EPVEKQGVDDIEKFEDGSGAGFESSETGDYEEKAEETAEAP 951
 QY 899 ---GSGRATTAKSGAQODSSSATCSADVEDEABGGDKNRLSPRPSLLTPTGPRANA 955
 DB 952 EEDGEDNAGSGASKHSPTEDDESAAEADV-----HLKKEKRESV--SGDDRAEE 999
 QY 956 SPQKPLDLKQLKQRAAAIPIQVTKVHEPPREDAAPTAPAPPPPPQNLQPSD----- 1010
 DB 1000 DMDDVLEGEAFQS-----EEGEEDKADAREEGYEPDKTAEDYVMAVADKAAEA 1052
 QY 1011 -----APOQPG--SSPRGKSRSPAPP-----ADKEAFAAEAKLPDGPDC 1048
 DB 1053 GVTERQYGLTSAKQPGIQSP---SREPASSIHDETLPGGSESEATASDENREDQPEE 1109
 QY 1049 WTSGLPFPVPREVTKASHAPDPSAFYAPGHPPLPLGLHDTARPVLP-----RPRTSN 1104
 DB 1110 FTATSGYQSTIEI-----SSEPTMDMSTPRDVMSETNNETESP 1152
 QY 1105 PPPLTSSAKHPSVLERO-----ICAIQSGMSVOLHPVSEHAKAPVGVPTMGLPLP 1155
 DB 1153 SQEFNITKYESSLYSQEYKPAVASFNGLSGSKTDTATDKDYNASAS-----TISPSS 1208
 QY 1156 MDPKKLA-----PFGSVKQEOQLSPRGQAGPPESGLGYPTAQEASV 1194
 DB 1209 MEEDKFSKALRDAYCSEKELKASAEIDIKDVSDERLSP--AKSPSLSPSP-----SP 1261
 QY 1195 LRGTALGVPGSGITKGPSTRVPSDSAITYRGSIHTGTPADVLKYGITRIIGEDSPSR 1254
 DB 1262 IEKTPIGE-----RSVNFSLTPNEIKVSAEARSVSP-----GVTQAVVEHCASP 1308
 QY 1255 LDRGREDSLPKGHVYIEGKGHVLSYEGGMSVTCQSKEDGRSSSGP-----PHETAAPKRT 1310
 DB 1309 BEKTLVVSPSQSV--TGSAGHTPYQ-----SPTDEKSSHLPTVESENAQAVP--- 1355
 QY 1311 YDMBGRVGRAISSASIEGLMGRAP-----PERHSPHLKEQHHRGSIQTGIPRSVVE 1365
 DB 1356 -----VSFESEAKDENERSLSPMDPEVPDSESPVE--KVLSPLRPLILGSESPY-- 1405
 QY 1366 AQEDYLREAKLLR-----EGTPPPPPPSRDLTEAYKTAQALGPLKPAHEG 1413
 DB 1406 --EDFLSADSKVLGRRSSESPFEGKNGKQGFDPRESVPVSDLT-----STGLYQDKQBEKSTG 1459
 QY 1414 LVATVYAGRSITHEIPREELRHTPELPLAPRLKSGSITQGTPLKYDTGASTGSKKHIDV 1473
 DB 1460 FIPKEDGPEKKTSDVETMSSQSALADERKL--GGDV---SPTQID--VSQFGSKEDT 1513
 QY 1474 RSLIGSPGRTFPVHPLDVMADARALERACEESLKSRRPTASSSGGSIARGAPVIVPEL 1533

```
Db 1514 KMSISEGTVSDKSAFPVD-----EGVAEDTYSHMEGVASVSTASVATSS---FPPEP 1561
Qy 1534 GKPROSLTYEDHGAPEAGHLPRGSPVTMMREPTPLRQEGSLSSSKASQDKLTSTPREI- 1592
Db 1562 TDDVSPSLHAEGVSGHPSTVEDVDSLSVVQPTTFQETEMSPSKBECPRPMISPPDFES 1621
Qy 1593 ----AKSPHSTVPEHHPHPISPYEHLLRGVSGVDLYRSHIPFLAFDPTSPRGIPLDAAAY 1649
Db 1622 PKTAKS-RTPVQDHRSE-----QSSMSIEFGQESPHSPAMDFS--- 1659
Qy 1650 YLPRHLANPTPHLYPPYLRGYPDTALENRQIINDYITTSQQMHHNTATAMAQRADM 1709
Db 1660 -----RQSPDPTL-----GASVLHITENGTEV-DY----- 1685
Qy 1710 LRGLSP-----RESSALNVAAGPRGIIDLSQVPLPVLVPTPGTATAMDRLAYLPTAPO 1766
Db 1686 ----SPCIDQSSLSHK-----IPTEPSYTDNDLSELISVSQ 1721
Qy 1767 -----PFGSRHS-----SSPLSGGTHLTKPTTTS-----SSERDRDRDRDREREK 1812
Db 1722 VEASPTSSAHTPSQIASPLQEDTLSDVVPVPPREMSLYASLASEKVQSLGE-----KLSPK 1777
Qy 1813 SLITTTTVEHAPIWRPGTEQSGSGSGSGGGGSSRRPASHAHQHSPIPRTDALQ 1872
Db 1778 SDISPLTPRESSPLXPCGFSSTSAKETAAAHQASSP-----PIDAATAEPTG 1827
Qy 1873 QRPVSLHNTGMGIITAVEPSKPTVLRSTSTSSVPRPAATFPATHCPGLGTLGDIYPTL 1932
Db 1828 FRSSMLFDMQHHL-----ALNRDLTSSVEKDSGCKTP-----GDFNAYQKP 1871
Qy 1933 MBPVLLPKPEA-----PRVARPERPRADTGHAF-----LAKPPAR 1966
Db 1872 ENAAGSPDEEDVDYESQEKTIHWDVRYVEKTEKTIKPCDSGVSYETIEKTKTPED 1931
Qy 1967 SGL-----EPASSSKG-----SEPRPLVPVPGHATIAATPAKNLAPH----- 2005
Db 1932 GGYTCETEKTTRTEGGYSYEISEKTRTPEVSGYTYEKTERSRRLDDLSNGYDDTE 1991
Qy 2006 ---HASDPAPPASADPHREKTSQKPSIQEELERSLGYH-----GSSVSPSG 2052
Db 1992 DGGHILGD-----CSVSYETTEKISFP-----ESESYSYTSYKTRSPDTSAYCYET 2040
Qy 2053 VFPVSPVSSPS-----LTHDKGLPKH-----LEELDKSHLEGELRPK--QP 2091
Db 2041 MEKITKTPQASTYSYETSDRCYTTEKKSPEARQVDVLCVSSCEFKHPKTELSPSFINP 2100
Qy 2092 GPVK-LGGEAAHLPHLRPLPESQSPSSPLLQT--APGVKGHORVVTVAQHISEVITQDYT 2148
Db 2101 NPLEWFAGE-----EPTSESE--KPLTQSGGAPPPSGGKQ---QGRQCDETPTPTSVS 2147
Qy 2149 RHHPQOLSAPLAPLYSFGASCPLV-----DLRRPPSDLYL-----PPDHGA 2192
Db 2148 ESAPQSDSDVPET-----EBCPSITADANIDSEDESETIPTDKTVTVYKHMDDPPP--A 2199
Qy 2193 PARGPSHSGGKRSPPNKTSVLGGGEDGI-----EPVSPPEGMTEPFGHSRAV 2241
Db 2200 PMQ-----DRSPSPRIPDVSMWDDPALAVDQNLGKAVKDKLEKTKTKPKTKKS 2251
Qy 2242 YPLLVRDQSGTEPSRMGSKSPGNTSOPPAFFSKLTESNAMYKSKQKQENKLNTHNRNE 2301
Db 2252 SPVKKGDKG-----SKPLAASPCKGALKESDKVSRVASPKKESVEKA-TKTTTT 2301
Qy 2302 PEYNTISQPTGEIFNMPAITGTLMYRSQAVQEHASTNNGLEAIRKALMGKYDQWEEESP 2361
Db 2302 PEY-----KATRG-----EEDK 2313
Qy 2362 PLSANAFNPLNASLPAAMPIITADGRSDHTLTSPGGGKAKGVSRPSRKAAPGL 2421
Db 2314 KETKNA--ANASAKSAK--TAT-----TGPGTTKTAKSIVP-----PGL 2350
```

RESULT 11

```
YCS3_YEAST
ID YCS3_YEAST STANDARD; PRT; 1226 AA.
AC P25357;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 138.5 KDA PROTEIN IN RPS14A-GNS1 INTERGENIC REGION.
GN YCR033W OR YCR33W OR YCR592.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92133166; PubMed=1776366;
RA Wicksteed B.L., Roberts A.B., Sagliocco F.A., Brown A.J.P.;
RT "The complete sequence of a 7.5 kb region of chromosome III from
RT Saccharomyces cerevisiae that lies between CRY1 and MAT.";
RL Yeast 7:761-772(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91335897; PubMed=1872032;
RA Jia Y., Slonimski P., Herbert C.J.;
RT "The complete sequence of the unit YCR59, situated between CRY1 and
RT MAT, reveals two long open reading frames, which cover 91% of the
RT 10.1 kb segment.";
RL Yeast 7:413-424(1991).
CC -!- MISCELLANEOUS: THIS PROTEIN IS ENCODED BY A NON-ESSENTIAL GENE.
CC -!- SIMILARITY: SOME, TO S.POMBE SPAC2E12.01.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59075; CAA41799.1; -
DR EMBL; X59720; CAA42300.1; -
DR PIR; S15053; S15053.
DR PIR; S19445; S19445.
DR SGD; S0000629; YCR033W.
DR InterPro; IPR001005; -
DR Pfam; PF00249; myb_DNA-binding; 2.
DR PROSITE; PS50090; MYB_3; 1.
KW Hypothetical protein.
FT DOMAIN 594 599 POLY-SER.
FT VARIANT 305 305 T -> A.
FT VARIANT 375 375 N -> S.
FT VARIANT 404 404 H -> Q.
FT VARIANT 435 435 A -> V.
FT VARIANT 442 442 D -> N.
FT VARIANT 644 644 Y -> H.
FT VARIANT 801 801 E -> D.
FT VARIANT 806 806 K -> Q.
FT VARIANT 831 831 I -> V.
FT VARIANT 846 846 E -> G.
FT VARIANT 851 851 A -> D.
FT VARIANT 855 857 GVR -> AVQ.
SQ SEQUENCE 1226 AA; 138503 MW; 8D133A0918658E53 CRC64;
```

Query Match 3.1%; Score 412.5; DB 1; Length 1226;

Best Local Similarity 19.5%; Pred. No. 4.3e-07;

Matches 240; Conservative 177; Mismatches 439; Indels 377; Gaps 47;

Qy 18 RYPHPSLSYPVQ-----IARTHTDVGLEY 42

Db 85 RYDPSPVSRPSSSYSTRKIGSRYPNDYVERSSSTTSSTPESMNTSTIHTNTDIGNRY 144

Qy 43 QH-----HSRDYASHLSPGSTIQPQRRRSLSEFPQG-----NER 78

QY 1903 TSSVPRPAATFPATHCPLGGTLDGVYPTLMEPVLLPKEA----- 1942
Db 1862 GYTP-----GDFSAYAKPETTRSPDEEDYDSEYKTKTRTSDVGGY 1905
QY 1943 -PRVAPERPRADTGHAFF--LAKPARSGLEPASPSKSGSEPRPLVPVSGHATARTPA 1999
Db 1906 YEKIIRTTKSPSDSGSYETIGK-----TTKTPEDGDYSYEIEKTT-----RTPE 1951
QY 2000 KNLAPHASPDPPAPPASADPHREKTQSPKPSIQEILRSYHGSSYSPGCVFVSPV 2059
Db 1952 EGGYSYDISEKTTSP-EGVSGYSYEKTERSRLLDDI---SNGYDDSE---DG----- 1997
QY 2060 SSPSLTHDKGLPKLEELDKSHLEGELRPKQPKVLGGEEAHLPLRLPESQSSPL 2119
Db 1998 -----GHTLGDPSYETEK-----ISFPSEGS----- 2024
QY 2120 LGTAPGVGHORVVTLAQHISEVITQDYTRHPQOLSAPLPAPLYSFPGASCPVLDLRP 2179
Db 2025 YETSVKTRTPDTSTYCYETAETKTRT-----PQ-----ASTYSYETSDLCYTAEKKS 2072
QY 2180 PSDL-----YLPDPDHGAPARGSPHSEGGKRSPEPNKTSVLGGGEDGIEPVS 2226
Db 2073 PSEARQDVLCVSCSEYKHPKTELSPSPINPNLEWFASEEPTSESEKPLTQSGGAP-- 2130
QY 2227 PPEGMTPEGHSRAVYPLLYRDGEOTEPSRMGSKSPGNTSQ--PPAFTSKLTESNSAMVK 2284
Db 2131 PPGGKQGG-----RODETPPISVESAPQSDVPPE-----TE----- 2167
QY 2285 SKQOEINKLTHNNEPPYNSIQPTEIFNMPATITGTLMTYRSQ-----AVQE----- 2334
Db 2168 -ECPSITADANIDSEDESE-----TIPTDKTVTYKHMDDPPAPVQDRSPSP 2212
QY 2335 -HASNW-----GLEAIRKALMGKYDOWEESPLSANFNPINASASLPAAMPITAAD 2387
Db 2213 RHPDVMVDPEALAEQNLGKAL--KKDLKKT-----KTKPGTKTKSSSPVKKSD 2262
QY 2388 GRSDHTLTPSGGGGKAKVS--GRPSSRAKSPAGLSDGDRPPSVSVHSEGCNCRRT 2444
Db 2263 GSKSLAASPKPAGLKESDKVSRVASPKKESVEKAAPTTTPEVKAARGEKDEKTKN 2322
QY 2445 LNRVWEDRPSAGSTPPYPNPLMRLOAGVNASPPPGCLPAGSGPLAGPHHA 2497
Db 2323 AANASAKSATATAGP-----GTTKTKSSAVPPGLPVYLDLCYIPNHS 2367

RESULT 13
CAIL_HUMAN STANDARD; PRT: 1464 AA.
AC P02452; Q15176; Q14037;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I)
RT chain of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [2]
RP SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84270697; PubMed=6462220;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA Myers J., Williams C., Ramirez F.;
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary

RT conservation of a pattern of introns and exons.";
RL Nature 310:337-340(1984).
RN [3]
RP SEQUENCE OF 162-301.
RX TISSUE=Skin;
RC MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN [4]
RP SEQUENCE OF 263-268.
RX TISSUE=Skin;
RC MEDLINE=71001508; PubMed=4319110;
RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT "A comparative study of glycopeptides derived from selected
RT vertebrate collagens. A possible role of the carbohydrate in fibril
RT formation.";
RL J. Biol. Chem. 245:5042-5048(1970).
RN [5]
RP SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=5689127;
RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA Prockop D.J.;
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
RL of structures that are conserved during evolution.";
RL Biochemistry 22:5213-5223(1983).
RN [6]
RP SEQUENCE OF 1229-1454 FROM N.A.
RX TISSUE=Bone;
RC MEDLINE=88124208; PubMed=3340531;
RA Mackelae J.K., Raassina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT domain.";
RL Nucleic Acids Res. 16:349-349(1988).
RN [7]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88097389; PubMed=3480516;
RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT "Regulatory elements in the first intron contribute to
RT transcriptional control of the human alpha 1(I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN [8]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [9]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=88033098; PubMed=2822714;
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA de Wet W.J.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
RT collagen gene enhance transcription.";
RL J. Biol. Chem. 262:15151-15157(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).

RP REVIEW ON OI VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;
RA Byers P.H., Wallis G.A., Walling M.C.;
RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL J. Med. Genet. 28:433-442(1991).
[13]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=97169389; PubMed=9016532;
RA Dalgleish R.;
RT "The human type I collagen mutation database.";
RL Nucleic Acids Res. 25:181-187(1997).
[14]
RP VARIANT OI-II CYS-1166.
RX MEDLINE=86287390; PubMed=3016737;
RA Cohn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
change in one human pro alpha 1(I) collagen allele.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
[15]
RP VARIANT OI-II ARG-569.
RX MEDLINE=87222295; PubMed=3108247;
RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT "Lethal perinatal osteogenesis imperfecta due to the substitution of
arginine for glycine at residue 391 of the alpha 1(I) chain of type I
collagen.";
RL J. Biol. Chem. 262:7021-7027(1987).
[16]
RP VARIANT OI-II CYS-926.
RX MEDLINE=88033031; PubMed=3667599;
RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT "A point mutation in a type I procollagen gene converts glycine 748
of the alpha 1 chain to cysteine and destabilizes the triple helix in
a lethal variant of osteogenesis imperfecta.";
RL J. Biol. Chem. 262:14737-14744(1987).
[17]
RP VARIANT OI-II ARG-842.
RX MEDLINE=88298828; PubMed=3403550;
RA Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;
RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
chain in lethal perinatal osteogenesis imperfecta. Demonstration of
the peptide defect by in vitro expression of the mutant cDNA.";
RL J. Biol. Chem. 263:11627-11630(1988).
[18]
RP VARIANT OI CYS-1195.
RX MEDLINE=89218628; PubMed=3244312;
RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT "A cysteine for glycine substitution at position 1017 in an alpha
1(I) chain of type I collagen in a patient with mild dominantly
inherited osteogenesis imperfecta.";
RL Mol. Biol. Med. 5:197-207(1988).
[19]
RP VARIANT OI-II VAL-434.
RX MEDLINE=89255493; PubMed=2470760;
RA Patterson E., Smiley E., Bonadio J.;
RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
mutation.";
RL J. Biol. Chem. 264:10083-10087(1989).
[20]
RP VARIANT OI-IV SER-1010.
RX MEDLINE=89308591; PubMed=2745420;
RA Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koeplin D.A.;
RT "Osteogenesis imperfecta type IV. Detection of a point mutation in
one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
RL J. Biol. Chem. 264:11893-11900(1989).
[21]
RP VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.
RX MEDLINE=89380165; PubMed=2777764;
RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
RT "Characterization of point mutations in the collagen COL1A1 and
COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL J. Biol. Chem. 264:15809-15812(1989).
[22]

RP VARIANT OI SER-1022.
RX MEDLINE=90062068; PubMed=2511192;
RA Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
RT "Substitution of serine for alpha 1(I)-glycine 844 in a severe
variant of osteogenesis imperfecta minimally destabilizes the triple
helix of type I procollagen. The effects of glycine substitutions on
thermal stability are either position of amino acid specific.";
RL J. Biol. Chem. 264:19694-19699(1989).
[23]
RP VARIANT OI-II CYS-1082.
RX MEDLINE=89109573; PubMed=2913053;
RA Constantinou C.D., Nielsen K.B., Prockop D.J.;
RT "A lethal variant of osteogenesis imperfecta has a single base
mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
chain of type I procollagen. The asymptomatic mother has an
RT unidentified mutation producing an overmodified and unstable type I
procollagen.";
RL J. Clin. Invest. 83:574-584(1989).
[24]
RP VARIANT OI CYS-272; CYS-704 AND CYS-896.
RX MEDLINE=90009313; PubMed=2794057;
RA Starman B.J., Eyre D., Charbonneau H., Harrylock M., Weis M.A.,
Weiss L., Graham J.M., Byers P.H.;
RT "Osteogenesis imperfecta. The position of substitution for glycine by
cysteine in the triple helical domain of the pro alpha 1(I) chains of
type I collagen determines the clinical phenotype.";
RL J. Clin. Invest. 84:1206-1214(1989).
[25]
RP VARIANT OI-II CYS-422.
Query Match 3.0%; Score 400.5; DB 1; Length 1464;
Best Local Similarity 21.5%; Pred. No. 1.3e-06;
Matches 380; Conservative 102; Mismatches 642; Indels 647; Gaps 83;
QY 722 EALHAGNEVPREGCGPATVNNSSDTESIPSPHTAAK-DTQGNQPKPPA-TLGADGPP 779
DB 77 ETKNCPGAEVPEGEC-PVCPDGSESPDTQETTVGEGKGTGPRGPRGPPGPRDGIP 135
QY 780 -----PGPPTPRRTSRAPETTPASEANGATPP-----P 810
DB 136 GQGLPLGPPGPP-----GPPGPPGLGNFAPQLSYGDEKSTGGISVP 178
QY 811 APPPSAP-----PPVVPKEEKEETAAAPPVVEEG-----EOKPAAEELAYDTCK 857
DB 179 GPMGSPGPRGLPGPGAGPQGQFQ-----GPPGEPGEGASGPMGPRGPPGPKNGDDGE 234
QY 858 AEPVKSECTEAEAECPAKGDAEAAEAETGALKAEKKEGSGRATTAKSGCAPQSDS 917
DB 235 AGKPCR-----PCEKGGPPGQARGLPQTA--GLPMKMGHRGFGSLDGAKGDAGP----- 282
QY 918 SATCSADEVDEAEGDKNRLSP-----RPSLLTPTGDRANASPKPQLDLKQLKQRAAI 973
DB 283 -----AGPKGEGSPGENGAPQMGPRGLPGERGPPGAP-----GPAGAR 322
QY 974 PFIQVTKVHEPPREDAAATKPA-PPAPPPQNLPQSDAPQQPGSS--PRGKSRSPAPP- 1029
DB 323 GNDGATGAAGAPP-----GPTGAPGPPGPGAVGAKGEA-GPQGRGSEGGQVGRGEPGPG 377
QY 1030 -----ADKEAFAAEAKLPGDPPTWTSGLP-FPVPPREVIKASHPADPSAFS 1076
DB 378 PAGAGAPNAGDAGQPGAKGANGAPG-----IAGAPGPPGARGSPGPGPGPPGPKGN 432
QY 1077 YAPPGHPLPLGLHDYARPLYPRPPTISNPPPLITSSAKHPSVLERITGATISQGNMSVLHVP 1136
DB 433 SGEFAPGSGKG-DTGAKGEPGPVGQGP-----GPAGEEGKRG----- 470
QY 1137 YSEHAKAPVGPVTMGLPLPMDPKKLAPFSGVKQEQLSRQAGPESLGVPTAQAEASVLR 1196
DB 471 -----ARGECPPT--GLP-----GPPGERGGGSGRFPGADGAVGPK 505
QY 1197 GTA--LGSVPGGSITKGP-STRVPSDSAITVRSITHTGTPADLVLYKGTITRIIGSDPS 1253
DB 506 GPAGERGS-PGPAGPKGSGEAGRFEGALPGAKGLT-GSPGSPGPDGK-----TGPPGPA 559

DR EMBL; U84643; AAC51185.1; JOINED.
DR EMBL; U84644; AAC51185.1; JOINED.
DR EMBL; U84645; AAC51185.1; JOINED.
DR EMBL; U84646; AAC51185.1; JOINED.
DR EMBL; U84647; AAC51185.1; JOINED.
DR EMBL; U84648; AAC51185.1; JOINED.
DR EMBL; U84649; AAC51185.1; JOINED.
DR EMBL; U84650; AAC51185.1; JOINED.
DR EMBL; U84651; AAC51185.1; JOINED.
DR EMBL; U84652; AAC51185.1; JOINED.
DR EMBL; U84653; AAC51185.1; JOINED.
DR EMBL; U84654; AAC51185.1; JOINED.
DR EMBL; U84655; AAC51185.1; JOINED.
DR EMBL; U84656; AAC51185.1; JOINED.
DR EMBL; U84657; AAC51185.1; JOINED.
DR EMBL; U84658; AAC51185.1; JOINED.
DR EMBL; U84659; AAC51185.1; JOINED.
DR EMBL; U84660; AAC51185.1; JOINED.
DR EMBL; U84661; AAC51185.1; JOINED.
DR EMBL; U84662; AAC51185.1; JOINED.
DR EMBL; U84663; AAC51185.1; JOINED.
DR MIM; U84663; AAC51185.1; JOINED.
KW Disease mutation; Polymorphism.
FT DOMAIN 89 97 POLY-GLU.
FT DOMAIN 204 207 POLY-SER.
FT DOMAIN 616 619 POLY-SER.
FT DOMAIN 919 924 POLY-SER.
FT DOMAIN 1285 1289 POLY-LYS.
FT DOMAIN 1375 1386 POLY-LYS.
FT DOMAIN 1398 1405 POLY-LYS.
FT VARIANT 53 53 W -> R (IN TCS).
FT VARIANT 439 439 /FTID-VAR_005630.
FT VARIANT 810 810 /FTID-VAR_005631.
FT VARIANT 1313 1313 /FTID-VAR_005632.
FT VARIANT 1355 1355 A -> V.
FT VARIANT 1411 1411 /FTID-VAR_005633.
FT SEQUENCE 1411 AA; 144312 MW; 3880203D985C2699 CRC64; /FTID-VAR_005634.
D -> G.
Query Match 3.0%; Score 391.5; DB 1; Length 1411;
Best Local Similarity 20.4%; Pred. No. 2.4e-06;
Matches 364; Conservative 178; Mismatches 664; Indels 581; Gaps 76;
QY 656 YKRNQILQOHLKMEKERNARKKKAPAAASEEAFFPVVEDEMEA----- 707
DB 50 YTHWOOTSELGRKAEEDAALQAKTRVSDPSTSESE-----EEEAETAATKATPR 104
QY 708 -----SGVSGNE-----EEMVEEAALHASNEVPRGECGPAT-----VNNSSDTE 749
DB 105 LASTNSSVLGADLPMSMEKAEKTEKAGTGNMSPH-----PATGKTVANLLSGKSPRK 159
QY 750 SI-PSPHTEAAKDTQONGKPKPATLGADPPPGPPTPPR----- 787
DB 160 SAEPNATTLVSETEEGSV--AFGA-AAKPGMVSAGQADSSSEDTSSTSSDETVEVKA 216
QY 788 -----FTSRAPIEPTPASEATGATPPAPSPAPPPVVKKEKEETAAAPVVEG 840
DB 217 SEKILQVRAASAPAGTGPCKGAT-----PAPPGKAGAVASQTKAGKPEDESSESS 270
QY 841 E-EQPPRAEEL--AVDTCKAEPEPVKSECTEAEAGCPAKGDAEAAEATGALKAEKE 897
DB 271 DSEETPAKALLQAKASGKTSQ--VGAASAPAKESPRK-----AAPAP 314
QY 898 GSGRATTAKSGAPQDSSATCSADEVDEAGGDKNLLSPRSLTPTGDPANASP 957
DB 315 GTGFAVAKAAGKREDSQSSSESDSEEA-----PAQAKPSG----- 354
QY 958 OKPLDLKQLKRAAATPIQVTKVHEPPREDAATKPAAPPAPPPQNLQOPESDAFQQPGS 1017
DB 1411 AA; 144312 MW; 3880203D985C2699 CRC64; 1411

DB 355 -----KAPQVRAASAP-----AKESPRKGAAPA-----PPRKTGPAAAQVOVGKQ 394
QY 1018 SPRGKSRSPAPPADKEAF--EAQ-----KLPGDPCHTSGL--PF-----PVPPREVIK 1064
DB 395 EEDSRSSSEESDRELAAMNAQVKPLGKSPQVKGKPGPASTMGMPGLGAGVPVPGKVG 454
QY 1065 ASPHA-----PDPSAFSYAPPCHPLPLGLHDTARVLP RPPTISNPPPLISSAKHPSVL 1118
DB 455 ATPSAQVGKWEDESSES-----EESDSDSGEVPITAVAP-----AQ 491
QY 1119 ERQIGAISQMSVOLHVPYSEHAKAP--VGPVTMGLPL--PMDPKKLAPFSGVKOEQLS 1173
DB 492 EKSLGNILQAK-----PTSPAQPPQKAGPVAVQVKAEPMDNSE-----S 533
QY 1174 PRQAGPPESLGVPTAOEASVLRGALGSPVPGSITKGIPTSTRVPSDASITTYRGSTHGT 1233
DB 534 SEESDSDASEEAPAAWTAQAQ-----PALKIPQTKACPKKTNTTASA 577
QY 1234 PADVLYKGTITRIIGEDSPSRDLRGREDSLPKGVHYIEGKGHVLSYEGMSV-----TQC 1289
DB 578 KVAPVRVGT-----QPPR-----KAGTATSPAGSSPAPVAGGTQR 611
QY 1290 SKEDGRSSSGPHETAAPKRTYDMMEGR--VGRAT-----SSASIEGLMGRAPPRHSP 1342
DB 612 PAED--SSSSEESDEEKTGLAVTVGQAKSVGKGLQVKAASVPVKGSLGCGTAPV---- 665
QY 1343 HHLKEQHHRIGSITQIGIPRSYVEAQEDYLRREAKLLKRGSTPPPPSRDLTEAYKTOAL 1402
DB 666 -----LPGKTGPTVTVQKAEKQED-----SESSEESDEEAAAQAVKTSVKKTQA- 713
QY 1403 GPLKLPKPAHEGLVATVKEAGRSIHEIPRELRHLPRLAPRLKESITQCTPLKYDTG 1462
DB 714 ---KANPA-----AARAPSAKGTIS--APGKVVA 738
QY 1463 ASTGSKKHDRSLIGSGPRTPPVH-----PLDVMADARALERACTEESLSKR 1511
DB 739 AQAQAKOR-----SPSKVPPVRNPQNSTVLARGPASVPSVGKAVATAA-----QAQ 784
QY 1512 PGTASSSGSIIARGAPVIVPE-LGKPRQPLTYEDHG--APPAGHLPRGSPVPMREPTPR 1568
DB 785 TQPEEDSGSEESDESEEAETLAQAKPSGKTHOIRALAP-----AKESPRKGAAPTTP 839
QY 1569 LQEGSLSSSKASQDRKLTSTPREIAKSPHSTVPEHHHPHPIPSYEHLLRGVSGVDLYRSHI 1628
DB 840 GKTGPSAAQAGKQDSS--GSSSEESDSDGEAPAAVTSQAQVTKP----- 880
QY 1629 PLAF-DPTSI PRG---IPLDAAAAYLPHLAPNPYPHYLPYLLIRGYDPTAALNRQT 1684
DB 881 PLIFVDPNRSAPGAPATPAQAQAA-----STPRKARASESTARSSSES 924
QY 1685 IINDYITTSQ-----MHNTAT-----AMAOADMRLGLSPRESSLALN--YAAQPRGI 1731
DB 925 EDEDVTPATQCLTPTGRTVNTVMTPTAHPRIAPKASWAGASSKESRISDGKKQEGPATQ 984
QY 1732 IDLSQVPHLVLPVPTPGTPATAMDRLAYLTPAQPFSSR--HSSSPLSPGGTHLTKPT 1789
DB 985 VSKKNPASLPL-----TQAALKVLAQKASEAQPPVARTQPSGVDASVAGTLPATSPQ 1036
QY 1790 TTSSEERDRDRDRDREREKSIILTSITVVEHAPINWPGTEQSSGSSGSSGGGSSS 1849
DB 1037 STSVQAKGTNKLK-----KPKLPEVQQAATKAP--ESSDSDSDSDSSGSEEDGE 1084
QY 1850 RPASHSHAHQHPISPRTOTDALQORESVLHNTGMKGIITAVEPSKPTVLRSTSTSPVRP 1909
DB 1085 GPGKAKSAHTLGTPTSRTELVEE-----TAAESSEDDVV----- 1119
QY 1910 AATFPATPCPLGGTLDGVYPTLMEPVLLPKEAPRVARPERPRADTGHAFLAKPPARSGL 1969
DB 1120 -----APQSLLSG-----YMTPLTP--ANSQAKATPKLD-----SPSPVSSYL 1158
QY 1970 EPASSPSKSGSEPRP-----LVPPVSG--HATARTPAKNLAPHHSPDPAPPASADPH 2022
DB 1159 AAKDDPDGQKQAKPQAAGMLSPKGTGKEAASGTTTPOKRRKKPKKAGNPNQASTLALQSN 1218

Qy	2023	REKTOSKPPSIQIELELRSLGYHGSSYSPGVEPVPVSPSSPLTHDGKLPKHLBELDKSHL	2082
		: : : : : :	
Db	1219	TQCLLGQPWPLNEAQOYA--	1246
		: : : : : :	
Qy	2083	EGLRKPQCPVVLKGEEAAHLPRLPEPQSPPSPLLQTAPGVKGHORVVTLAQHISEV	2142
		: : : : : :	
Db	1247	EQE-RKKYDVTTKESSRKGWESKRKLSDQPA--	1290
		: : : : : :	
Qy	2143	ITODYTRHHHPQQLSAPLPALPYLFPGACPCVDLDRRPESDLYPPOHGPARGSPHSEG	2202
		: : : : : :	
Db	1291	--	1295
		: : : : : :	
Qy	2203	GKRSPEPNKTSVLGGEDGIEPVSPPEGMTPEGHRSASVPLYLYRDCEQTEPSR----	M 2257
		: : : : : :	
Db	1296	GEASVSPEKTISTTSGKAARDKAS---GDVKKKKGGS-----LGSQAKDDPEEELQKGM	1348
		: : : : : :	
Qy	2258	GKSKSPGNWTOPPAFFSKLFTESNASVMVKKSQEIN--KKLTNHNREP	2302
		: : : : : :	
Db	1349	GTVEGQQSNPKSKKEKKKSDDRKKDKDEKKKKKKAKASTKDSESP	1395
		: : : : : :	

```

RESULT 15
CALL_CANFA ID CALL_CANFA STANDARD; PRT; 1460 AA.
AC Q9XSJ7;
AC 30-MAY-2000 (Rel. 39, Created)
AC 30-MAY-2000 (Rel. 39, Last sequence update)
AC 30-MAY-2000 (Rel. 39, Last annotation update)
AC COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
AC COL1A1.
AC Canis familiaris (Dog).
AC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
AC NCBI_TaxID=9615;
AC [1]
AC SEQUENCE FROM N.A.
AC TISSUE=Skin;
AC Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
AC "Sequence of normal canine COL1A1 cDNA.";
AC Submitted (MAY-1999) to the EMBL/genbank/DBJ databases.
AC -I- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
AC (FIBRILLAR FORMING COLLAGEN).
AC -I- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
AC -I- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
AC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
AC -I- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
AC -----
AC This SWISS-PROT entry is copyright. It is produced through a collaboration
AC between the Swiss Institute of Bioinformatics and the EMBL outstation -
AC the European Bioinformatics Institute. There are no restrictions on its
AC use by non-profit institutions as long as its content is in no way
AC modified and this statement is not removed. Usage by and for commercial
AC entities requires a license agreement (See http://www.isb-sib.ch/announce/
AC or send an email to license@isb-sib.ch).
AC -----
AC EMBL: AF153062; AAD34619.1; -.
AC InterPro: IPR000885; -.
AC InterPro: IPR001007; -.
AC Pfam: PF01410; COLFI; 1.
AC PROSITE: PS01208; VWFC; 1.
AC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
AC Glycoprotein; Collagen; Signal.
AC SIGNAL 1 22 BY SIMILARITY.
AC PROPEP 23 157 AMINO-TERMINAL PROPEPTIDE.
AC CHAIN 158 1214 COLLAGEN ALPHA 1(I) CHAIN.
AC PROPEP 1215 1460 CARBOXYL-TERMINAL PROPEPTIDE.
AC DOMAIN 34 92 VWFC.
AC DOMAIN 158 174 NONHELICAL REGION (N-TERMINAL).
AC DOMAIN 175 1188 TRIPLE-HELICAL REGION.
AC DOMAIN 1189 1214 NONHELICAL REGION (C-TERMINAL).
AC SITE 741 793 CELL ATTACHMENT SITE (POTENTIAL).
AC SITE 1089 1091 CELL ATTACHMENT SITE (POTENTIAL).
AC FT SITE

```

```

Db 834 -----AGPPGAGTGGPIGNVAGPCKGARGAGP----- 866
QY 1632 FDPSTIPRIGPLDAAAYYLRHLAPNPTYPHYLYRGGYDPTAALENRQTIINDYIT 1691
Db 867 -----PGATGPGAGRVGP-----PGPS-GNAGPP-----GPPGAGKEGGKARGE--- 908
QY 1692 SOOMHNTATAMAQADML-----RGLSPRESSLALNYAAGPRGIIIDLQVPHL 1740
Db 909 -----TGAPRGEVGPFGPCPAGEKSGPGADGPAGAPGTGPGOGIAGQRCVVGL 959
QY 1741 PVLVPPT--PGTPATAMDRLAYLPTAPQPFSSRHSSSLSPGCGPHLTXTPTTSSSERER 1798
Db 960 PQORGERGPGGLPFSGE-----PKQGPSGTGERGPGPMGPPGLAGPPGESG----- 1009
QY 1799 DRDRERDREREKSLTITTTVEHAP IWRPGTEOSSGSSGSG-----GGGSSSRPAS 1853
Db 1010 -----REGS-----PGAESGPRDGSPPGKGDGETGPAGPPGA 1043
QY 1854 HSHAHQHSPISPRTODALQORPSVLHNTGMKGIITAVEPSKPTVLKSTSTSSPVRPAATF 1913
Db 1044 PGAPGAPGVGP-----AGKNG-----DRGETGPAGPAG-- 1072
QY 1914 PPATHCPLGGTLDGVYPTLMEVLLPKEAPRVARPERPRADTGH-----A 1958
Db 1073 -----PIG-----PVGARGPAGPQGRGDKGETGEQDRIKIGHRGFS 1110
QY 1959 FLAKPPARSGLEPASSPSKSGSEPRPLVPVPSGHATARTPAKNLAPHASPDPPAPPASA 2018
Db 1111 GLOGPPGPGSGEGGPGSAGSP-----AGPRGPPGSA 1143
QY 2019 SDPHREKTQSKPFSIQELELRSLGYHGSYSPEGVEPVSPVSPSLTHDKGLPKHLEELD 2078
Db 1144 GSPGRDGLNGLPGPI-----GPPGPRGTGDAGVPVPPGPG----- 1180
QY 2079 KSHLEGELRPKQPGVYKLGGEAAHLPHLRPLPESQPSPLLOTAPGVKQH----- 2130
Db 1181 -----PPGPPGPPSGGDFSLP-----Op-----PQEKAHDDGGRYRAD 1215
QY 2131 -----RVVTLAQHISEVITODYTRHHHPQOLSAPLPAPLYSFPGASCPVLDL 2176
Db 1216 DANVVRDRDLEVDTTLSLSQOIEIRSPESGRKNPAR-----TC--RDL 1258
QY 2177 RRPSPDLXLPDPDHGAPARGSPHSGKRSPEPNKTSVLGG-----GEDGIEPVSP 2227
Db 1259 KMCHSDW-----KSGEYWDPNQCCNLDAIKVFCNMETGETCVYPTOP 1301
QY 2228 PEGM-----TEPGHSRAVYPLLYRDGEQTEPSRMGSKSPGNTSQPPAFFSKLTESNSA 2281
Db 1302 QVAQKNWYISKNPKEKRHWYGESMTDGFQFEYG-----GGSDPADVAIQLT----- 1349
QY 2282 MVKSKKQEIKNKLNTNHRNEPEYNISQPG 2310
Db 1350 FLRLMSTEASQNIYHCKNSVAYMDQQTG 1378

```

Search completed: September 8, 2001, 14:39:31
Job time: 433 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 13:19:33 ; Search time 92.74 Seconds
(without alignments)
3590.810 Million cell updates

Title: US-09-522-753-5
 Perfect score: 13215
 Sequence: 1 MSGSTQLVAQWTRATERYP.....WDEPKPLLCQVETLSDSE 2517

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_16:*

- ```

1: sp.arched:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.todent:*
12: sp.unclassified:*
13: sp.vibrate:*
14: sp.virus:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query % |  | Length | DB | ID     | Description         |
|------------|-------|---------|--|--------|----|--------|---------------------|
|            |       | Match   |  |        |    |        |                     |
| 1          | 1272  | 9.6     |  | 1047   | 4  | Q9NSZ0 | Q9nsz0 homo sapien  |
| 2          | 1075  | 8.1     |  | 3469   | 5  | Q9U412 | Q9u412 drosophila   |
| 3          | 1060  | 8.0     |  | 3502   | 5  | Q9VRJ9 | Q9vyj9 drosophila   |
| 4          | 576   | 4.4     |  | 3576   | 11 | Q9QZW2 | Q9qzw2 mus musculus |
| 5          | 574   | 4.3     |  | 5327   | 5  | Q76891 | Q76891 drosophila   |
| 6          | 569.5 | 4.3     |  | 3261   | 4  | Q9Y556 | Q9y556 homo sapien  |
| 7          | 552   | 4.2     |  | 1966   | 5  | Q9NHX6 | Q9nhx6 drosophila   |
| 8          | 552   | 4.2     |  | 2649   | 3  | Q9P3J0 | Q9p3j0 neurospora   |
| 9          | 550   | 4.2     |  | 1963   | 5  | Q9VSK5 | Q9vsk5 drosophila   |
| 10         | 546.5 | 4.1     |  | 2157   | 4  | Q95875 | Q95875 homo sapien  |
| 11         | 533   | 4.0     |  | 3080   | 5  | Q9VRY3 | Q9vry3 drosophila   |
| 12         | 522.5 | 4.0     |  | 2752   | 4  | Q9UQ35 | Q9uq35 homo sapien  |
| 13         | 520   | 3.9     |  | 2187   | 11 | P70670 | P70670 mus musculus |
| 14         | 519   | 3.9     |  | 3536   | 5  | Q9VZ30 | Q9vz30 drosophila   |
| 15         | 517.5 | 3.9     |  | 2157   | 11 | Q9Z1R1 | Q9z1r1 mus musculus |
| 16         | 512.5 | 3.9     |  | 2703   | 5  | Q9VEG7 | Q9veg7 drosophila   |
| 17         | 508   | 3.8     |  | 5027   | 11 | Q63460 | Q63460 rattus norv  |
| 18         | 504   | 3.8     |  | 3942   | 11 | O88737 | O88737 mus musculus |
| 19         | 503.5 | 3.8     |  | 7962   | 4  | O10465 | O10465 homo sapien  |

## ALIGNMENTS

|        |                                                      |                            |
|--------|------------------------------------------------------|----------------------------|
| RESULT | 1                                                    |                            |
| ID     | Q9NSZ0                                               | PRELIMINARY; PRT; 1047 AA. |
| AC     | Q9NSZ0;                                              |                            |
| DT     | 01-OCT-2000 (TrEMBLrel. 15, Created)                 |                            |
| DT     | 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)    |                            |
| DT     | 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  |                            |
| DE     | DE HYPOTHETICAL 113.8 KDA PROTEIN (FRAGMENT).        |                            |
| GN     | KDFZP434M075.                                        |                            |
| OS     | Homo sapiens (Human).                                |                            |
| OC     | Eukaryota; Metazoa; Chordata; Vertebrata;            |                            |
| OC     | Mammalia; Eutheria; Primates; Catarrhini; Hominoidea |                            |
| OX     | NCBI_Taxid=9606;                                     |                            |
| RN     | [1]                                                  |                            |
| RP     | SEQUENCE FROM N. A.                                  |                            |
| RC     | Tissue=Testis;                                       |                            |
| RA     | Duesterheft A., Lauber J., Mewes H.W., Gassenhuber   |                            |
| RL     | Submitted (JAN-2000) to the EMBL/GenBank/DDBJ datab  |                            |
| DR     | EMBL; ALH37641; CABY/0854.1; -.                      |                            |
| KW     | Hypothetical protein.                                |                            |
| FT     | 1                                                    |                            |
| SO     | SEQUENCE                                             |                            |
| FO     | 1047 AA; 113817 MW; B5A1EDE938B7222A C               |                            |

|    | Query Match           | 9.6%                                                         | Score 1272;        | DB 4;       | Length 1047; |
|----|-----------------------|--------------------------------------------------------------|--------------------|-------------|--------------|
|    | Best Local Similarity | 35.0%;                                                       | Pred. No. 2.5e-56; |             |              |
|    | Matches 401;          | Conservative 149;                                            | Mismatches 393;    | Indels 202; | Gaps 54;     |
| Qy | 1470                  | KHDVRSLLIGSPGRFFPVHPPLDVNMAD-ARALERACYE-----ESLKSRPGTFASSSGS | GIA                | 1523        |              |
| Db | 7                     | KNVKSLLITGPSKLSRGMPLELIPENIKVYERCKYEDVKAGETVTRSRHTSVWSSG     | PSVL               | 66          |              |
| Qy | 1524                  | RGAPVIVPELCKPQSPILTYEDHCA----PFAGHLRGRSGPVTMRRETPPRLOEGSLSSK | 1578               |             |              |
| Db | 67                    | RST---LHEAPKAQLSGIITDTSARTPTVSQNTWGRGSPW-----MNTSDVTLS       | SNK                | 118         |              |
| Qy | 1579                  | A-SODRKLTSTPRE-----IAKSPHSTVPEHHHPIDISYEHLLRSGVGDLYRSHIPLAFD | 1633               |             |              |
| Db | 119                   | STNHERKSTLTPTQRESIPAKSPVGGVDVVSH--SFPDPHHRGSTAGEVYRSHLP      | THLD               | 176         |              |
| Qy | 1634                  | PTSPRGPIPLD-AAAAAYLLPRHLAPNPTYPHLYPPYLIRGVPDPTAALEN-ROT      | INDYIT             | 1691        |              |









1016 ETFFKYLQHPKFGAIAAIDRKSPODCVRYVYLSKKTENYKQLLRKSROTRSSRNP 1075  
498 OOOOOO-----OOOOOOPRPS--QEEKDEKEKE-----528  
1076 KAQAOPOCIIDSMTGVTWMTLREQOOQSGRSVAERAEARAERVAEKAADA 1135  
529 ---AEKEEKEPE-----537  
1136 AKAESRAEKASATKAVEATAAGEKVAKAAAAAAAATATTATTITSSSTSSSSSA 1195  
538 -----VENDKEDLLK-----EKTDOTSG-----555  
1196 SSASTASSSTASPATLAGIAADKTDAGKTASAKNAATAGPTATGTPTAATPATATA 1255  
556 -----EDNDEKAVASKRKTANSOGRRKGRITRSMANEANSEALTPQOSA--602  
1256 PPEISAGEAKSKNAEEAATAAGATVATAGTPATGASASAGATATGATATAAAG 1315  
603 ----ELASMELNESSRWTEEE-----METAKKGLLEHGRNWSAIARMVSGSKTVSOCKN 651  
1316 VGKPEATAPAGTAAKGADSRPDANDPLAKTASAKAINAEGYN-----AIGNSSSSSSN 1369  
652 -----FFYNTKKRONLDE-----ILQOHK 670  
1370 ATGASAPVQVTLNGFKPGYQTVVMANYKASTGGDDSGANAGGAAPGSLAATNASIATSG 1429  
671 LKMEKERNARRKKKAPAAAEAEAFPPVVEDEMEASGV-----SGN-----713  
1430 DKIVKTTSSR-----APNSTSTAA-----NESSGAGVNYGHTATAGNYLGOKLKA 1479  
714 -EEMVEBAEALHAGNEVPR-----GECGPAT--VNNSSDPTESIPSPHTEAAKDTG 763  
1480 AQVEGLGAGNELHSDVSESKRKPELNSGEAGNATSAMTNSSTSGSMNISNHLKANA 1539  
764 QNG-----PPPPAT-----772  
1540 KDGSMMAKTSMASTSSASVVVTSTPSASSSSLSASSMLLISAASVMSTAAGATSSSTAT 1599  
773 -----LGADGPPPG-----PPTPRTSRA--792  
1600 TTATASATSLPLADGSGNSVWANEIILADGDKKLASCFCVCKAECPRTRPLKKGRCQO 1659  
793 ---PIEPTPAS-----EATGAPTPPPAPSPSPAPPV-----VPKEEKEETAAA 834  
1660 YGIPDETIPAGARVCNCSOCKSVRSRYPNCPLTPCPNKDRAQLRNIPSLFE---LA 1715  
835 PVVEE---GEEQKPPAA-----BELAVDTGKA-----EPPVKSE 865  
1716 PEVRDPLMAEFQIPPHATRCCSACLMIIRKLDLPQLNLTGSGSGAGSGGDETDVSTS 1775  
866 CTEEAEEGPAGKDAEAAEATAEGALKAEK-----895  
1776 SCDEREPG---GSDTASVESPEN---LQRHKSUTVWKVQOQOQOQOQOQOQOQOQOQ 1830  
896 -----KEGSGGRATTA-----906  
1831 SQOPPPPPAQOQKSGSGRGDGTPLIITPTRMSSKSGSGGAQTAGDNERLLPPAAGQA 1890  
907 -KSSGAPODSSATCSADEVDE-----AEGDKNRLLSRPR---SL 944  
1891 PKOKTSEYDSSATETADEENENSPANRQSPKVLFGHGHGHGHGHANNAVAGLOPPVANM 1950  
945 LTPGDPDRANASPOK---PLDLKQLKQRAAIPPIQ---VTKVHEPPREDAATPKPAPAP 999  
1951 GTGGVQVQGAAGQVINGPISM-----RREAVNNVQDCVFVIERSLKHGP-----1997  
1000 PPPQNLQPSDAPQOPG-----SSPRGKSRSPAPADKEAFAAQAQLPLGPPPCWT 1050  
1998 -----QPKGGGQOOGOGOGOGOGGTGPGOSGSPSOQOQOQOQOQOQOQOQOQOQ 2043  
1051 SGLPPFPVPPREVITKASHAPDPSAFSY-----APP-----GHPLPLGLHDTARPVLP 1098

2044 -----ERKELTIVREYRQDPGILKQOQOQOQOQAGGAPPTSAAAGSLPHG--TSVQKLTR 2094  
1099 PPTISNPPLISSAKHPSVLERQIGALISOGMSVOLHVPVSEHAKAPVGVTHGLPLMDP 1158  
2095 FAAVAPPPP-----AHP-LTPTSTCAGSNNGTSDSLATLSVVNSHMGVGIHGPMAH 2148  
1159 KKLAPFSVGKQOLSP--RGQAQPPPESLGVPITAOEASVL-RGTALGSPVPGSITTKGIPST 1215  
2149 ASSAGGIGVDKATITPVVKSSSGSGSSKSSASHSTATPTETIYNVPAHQPGICPP- 2207  
1216 RVPDSAI--TYRGSITHTPADVLYKGTITRI-IGEDSPSLDRGREDLSLPGHVIYEG 1272  
2208 --PSQSHVPAHPSHQTPAHFQSHSHGHTQLOYPEPQTLDLSIKKPPRDGSHPTG 2265  
1273 KKGHVLSEGGMSVTCSEKEDGRSS--GPPHETAAPK---RTYDMMEGRVGRAISSA 1325  
2266 -----AGSSSSGSGSGPSSDRHHGPPPTMSMKHIVRSGGMVGTGTVVPSLA 2316  
1326 -----SIEGLMRAIPPERHSPHHLKEOHIRGSITQIGIPRSYVEAO-----ED 1369  
2317 APSSYLYPTRSVKTIIGGGVWP-----GVLPVPCGSAlyLQVPVPVPI 2360  
1370 YLRREAKLLKREGTPPP--PPPSRDLTEAYKTOALGPLKLP--AHEGLVATVKEGRSI 1425  
2361 SISGQGLPPRAGQPPPAQPPSGRGVAKY-----PPKLSPOQAAHHLSPSHGHSPOQ 2413  
1426 HEIPREELRHTPELPLAPRPL--KEGSITQGTPLKYDTGASTGSKKHVDVRLIGSPGT 1483  
2414 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 2469  
1484 FPPVHPLDMDADALERACYEESILKS-RPGTASSG--GSTARGAPVIVPELGPXR--Q 1538  
2470 LSPKFD-----GLVRQTTPEGVSGVPGCAGSGKGHSITQGTPLHMP---PHHLE 2517  
1539 SPLTYEDHCAFPAGHLPRGSPVTMEPTPRLEQESLSSSKASQDRKLTSTPREIAKSPHS 1598  
2518 SKRPSEYKSKQSRHSPAQPGGNOOLPPPOOSS-----2552  
1599 TVPEHHPHPISPYEHLRGVSGVDLYRSHIPLAFDPTSPRGIPLDAAAAYYLPRHLAPN 1658  
2553 --PQAPP-----PQGYGVGVSSPY-----ARS 2572  
1659 PTPHYLYPYLIRGYPDTAALENROTIINDYITTSOOM---HHTATAMAQADMLRGL-- 1713  
2573 PFAGVVEQPOV-----LSTROIVMHDYITSOOMQOQOQOQOQOQOQOQOQOQOQOQOQ 2623  
1714 -----SPRES-SIALNYAAG-----PRG-----1730  
2624 SGGGGGGGSDKESPRNSVGSAGFAYGGDKESAPRGREYSRASPADHYNSGHDALA 2683  
1731 -IIDLS-QVPHLPVLVP-----PTGTPATAMDRL--AYLTPAPOFFSSRHSPLSPGG 1781  
2684 SFVDVAVQOQPLP--VPQOKDDKSPG-FSTAPGVGSGPPLGSPPLPHAVVGVQAQPPP 2740  
1782 PT-----HLT-----1786  
2741 PTAHHDQRYRLTLHHHHTLVQOQIAQQHYRSLNVAQVDMQMDQAKRVMRHOHQ 2800  
1787 -----KPTTSSSERDRDRDRDRDRERERERERERERERERERERERERERERERER 1813  
2801 VOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 2860  
1814 ILTSTTV-----EHAPITWPCTEOSSG--SGSSGGGGSSSSRSPASHAHQHSPI 1866  
2861 EODRARRVAREERHDS--RMRMFAGNVVTVGSGAGGGGPS-PGQFLRA-----2908  
1867 TODALQORPSVLHMTGKIITAVEPSKPTVLRSTSTSSPVPAATFPFATHCPCLGGTLD 1926  
2909 -----SVPETGPPRSIPDRERESYRQAHGGPAPEDTP--GQLS 2945  
1927 GUYPTLMEFVL---LPKEAPRVARPER--PRADTGHAFIAPKPPARS---GLEPASSPSK 1977  
2946 A--QSLIDAIITKHEINRSDATAGPREFPSPFVHAPL--PPRGSGGGGTGTTRSSPAN 3001



QY 734 GBCSGPATVNNSSDTEISPSHT-----EAKDTCQNGKPKPATILGADG-----PPPG 781  
D 1540 MOOKEKEKDQKPEAKEOBEPETHPKTPEAAETKEPEKAPVSAGLPAVITVVTPEPA 1599  
QY 782 PTP-----PRTSRAPTEPTPASEATCAPTPPPA-----PPSPSAPPVVPKEEKE 828  
D 1600 SSAPEKAEAEAPSPAGEKPAEPAPVSEETKLVSVPVVEQPROSDVPPGEDSRDQ 1659  
QY 829 EATAAP-----PVEEGEOKPAAAELEA-----V 853  
D 1660 DSAALAPSAQESAAATDAVPCVNAEPLTGTTVSQVSESVDPKPSQPQLSKLTORSEA 1719  
QY 854 DTGKAEPVKSCTE-----BAEOPAGKGDAAEAETAEGALKAEKGGSGRAT 904  
D 1720 EBGKVEKPDTPTEPDATONAGVASEAQAPASEDEANPPVA-----AKDRKTNKSRK 1775  
QY 905 TAKSSGAPODSSATCSADEVD-----EAE-----GG 932  
D 1776 TSVQAAASVVEKPVTRKSERIDREKLKRSSSPRGEAQKLEKMEAEKITRTASKSSGG 1835  
QY 933 DKNRLSPRLTPTGDPANASQKPLDLKQKORAAAIPIOVTKVHEPPREDAAPT 992  
D 1836 DTEH-----PEPSL-----PLSRRRNRVSVATMTDHSRSPAKEPVEQPVYTRKRLERLOE 1890  
QY 993 KPAPPAP-----PPONLOPESDAP-----QOPGSSPRGSRSPAPPADKEFAA-----1037  
D 1891 AVVPPTTPRRGRPPKTRRAEEDGHERKEPAETPRPABGWSRPSQKSAAGAAPOGKRG 1950  
QY 1038 -----EAKLGDPCPWTSGLPFPVPPREVI-KASHPADPSAFSAPPHPPLGLLHD 1090  
D 1951 RNEQVAAEAGAQAOSTREGNPKRGEREAASEPKRDRRDPSTDKSGDPTFPV-----2004  
QY 1091 TARPVLPPTTISNPPPLISSAKHPSVLERQJGALSQGSVQLHVPYSEHAKAPVGPVTM 1150  
D 2005 -----EVLERP-----PEKYKSRGARSTR-----SGMDRAAHORSLEMAAAGQAA- 2050  
QY 1151 GLPLPMDPKKLAPFGVQKQOLSP-RGQAGPPESLGVT-----AQEASVLRGTA-----1199  
D 2051 -----DKEAGPAAASQSESPQKSGSSPOLANNPADPDREAEESESASTAPPEGT 2103  
QY 1200 -----LGSVPGGSIITGIIPSTRVPSDAITYRGSITHG-TPA-----DVLK 1240  
D 2104 QLAROIELOAVONIAKLEPESAAAASKGTATATATASEEPAPHEGHKHPAQASETELA 2163  
QY 1241 GTITRIIGDPSRLDRGREDSLPKGHVIEGKGHVLSEYEGMSVTQCSKEDGRSSGP 1300  
D 2164 AAGSIISDASGEFNSAPPVPPGSGQTHPRE-----GMEPGLHEAESGILET 2214  
QY 1301 PHETAAPKRTYDMGRVGRAISSAIEGLMGRAIIPPERHSPHLKEQHHRIGSITQGP 1360  
D 2215 ATESAPQ-----VSALD-----PPEGSA-----DTKETRNGSDSV- 2246  
QY 1361 RSYVBAQEDYLREAKLKRGTTPPPPPSRD-----LTAAYKTOALGP 1404  
D 2247 -----QEAQSKAEYTPPRKDGROKTRRRNRANKKVVAITETRASEAQT 2294  
QY 1405 LKLKPAHEGLVATVKEAGRSIHEIPREELRHP-----ELPLAPRLKEGSITQGP 1456  
D 2295 QSESFAAEATATPEAQ--BEKSEKPPSPAECTFDPSPTPAESLQSENSAAEKTP 2352  
QY 1457 LKYDTGASTGSKKHDRSLIGSPGRTPPVHPDLD--VNADARALERACYEESLKSROP 1514  
D 2353 CK-----APVLPALPPLSQPALMDGDPQARFKVSHIESDPVT 2390  
QY 1515 ASSSG-----SIARGAPVIVPELKGKPROSP-----LTEDHGAPFAGHLPRGS 1558  
D 2391 PPSDSGIPPTTIPLVIAKLPVPIPG-GVPHQSPPPKVTENITROE-----E 2437  
QY 1559 PVTMRPTPRQEGSISS--SKASODRKLSTPREIAKS-----PHST--V 1600  
D 2438 PRAQSTPSPALPDPDKASMDTSSSTLRKILMDPKYVSATGVTSTSVTTAIEPVSAPCL 2497

RESULT 5  
O76891  
ID O76891  
AC O76891;

PRELIMINARY; PRT; 5327 AA.

QY 1601 PEHHHPHISPYPHLLRGVS-----GVDLYRSHIPLAEDFTSPRGIPLDAAAAAYLPRHL 1655  
D 2498 QEAPAPPCDPKHPLEGVSAAPVNPADTQASEVPVAADKEV---APV-----I 2543  
QY 1656 ANPTYPHYLPYLIRGYDPTAALENROFIINDYITTSQOMHHNTATAMAQRADMLRGLSP 1715  
D 2544 APKIT-----SVISRMVPSIDLENSQKI-----TLAKPAPQTLTGL-- 2579  
QY 1716 RESSLALNAAAGPRGIIDLSQVPHLPVLPVPTGTATAMRLAYLPTAPQPPSSRHSS 1775  
D 2580 -----VSALTGLVNVSLVP-VNALKGPVKSATLKLVS----- 2613  
QY 1776 PLSPGPTHLT-K-PTTSSSERERDRDRDREREKESILTSTT-----TVSHAPIWRP 1829  
D 2614 --TPAGPVNLLKGPVNVLTG-----PVNVLTPVSATVGTVAAP--GP 2653  
QY 1830 -----GTEOSSSSSGSGGGSSRPASHSHAHOSPISPRTQDQALQORPSVLHNTGMK 1884  
D 2654 VTAACGVATTGTAAVTGAVTAPAAKQKORASSNENSRFHGMSVIDDRPA---DTG-S 2709  
QY 1885 GIITAVEPSKPTVLRSTS---TSSPVRPAA---TFPPATHCPLGGTLDGYIPTLMEPVLL 1938  
D 2710 GAGLRVNTSEGVVLLSYSGKTEGQRIISAKISOIPASAMDI----- 2752  
QY 1939 PKEAPRVARPERPRADTGHAFKAPLAKPARSGLEBPASSPKSGSEPRPLVPVPGHATIARTP 1998  
D 2753 --EFQOSVKSQVKADS-----ITPTQSAKPGQPTPSAFANVAHSTLVLT- 2796  
QY 1999 AKNLAPHASPDPPAPPASASDPHREKTOSEFISQLELSRLGYHSGSYSPGVEPVPSP 2058  
D 2797 -----AOTYN-----ASP 2804  
QY 2059 VSPSLTHDKGLPKHLELDKSHLEGELRPKOPGPVKLGEEAAHLPHLRPLPESQSPSSP 2118  
D 2805 VLS-SVKTR---PSLEKPEPHLSVSTPYTOGGTVKVLTOGINTPPV--LVHNLVLTP 2858  
QY 2119 LLQTAAPGVKHQVRVYTLAQHISEVIT-----QDYTRHHPOOLSAPLPAPLY 2164  
D 2859 -----SIVTNKKLADPVLTKIETKVLOPANLGPILTTPHPPALPSKLPAEVN 2906  
QY 2165 SFP-GASCP---VLDLRRPPSDLYLPPDHGAPA---RGSHPSEGGKRSPEPNKTSVLG 2216  
D 2907 HVPSPSTPADRTIAHLATPKPDTHSPRPTGPTGTLFPRPCHPSPTTALSTNATVMLA 2966  
QY 2217 GG---EDGTEPSPBPTEPGHS-----RS 2239  
D 2967 AGIPVPOFISSIHPEQSVIMPPHSITQTVSLHLSQGEVMSPTPLPSITYSIRPETLHS 3026  
QY 2240 AVYPLLYRDGEQTEPSRMGSKSPGNTSQPPAFFSKLTESN-----SAMVSKKOE 2289  
D 3027 PRAPLQPOQIEARAPORVGTQPATTGVPALATQHPPEEVHVLVVARAAAPVQSEVLV 3086  
QY 2290 INKKLNTHNRPEYNI SOPTEIFNMPAITGTGLMTYRSQAOVEHASTNMGLEAIRKA 2349  
D 3087 MOSEYRLHPTVTPR-----DVRIMVHPHTAVSEQPRATEGVKVPVPANKAPQQLVKEA 3140  
QY 2350 LMKGKYDOWEESPLSANAFNPLNASASLPAAMPITAADGRSDHTLTPSGGGGKAKY-SGR 2408  
D 3141 -----VKTSIDAKAVPAPVAPVVPVPT-----PAPPHGEARILTVT 3178  
QY 2409 PSSRKAKSPAGLASGRDP-----SVSSVHSEGDGNNRRTPLTN-RWEDRPSSAGSTPF 2462  
D 3179 PSSQLQGLPL-----TPPVVVTHGVQIVHSSGELFQEVRYGVDVRYTHAPAQQLTHTQF 3231  
QY 2463 PY-NPLIMRLQAGVMSPPPPGLPAGSGPLAGP 2494  
D 3232 PVASSISLASRTKTSAQVPPPEGLQSTQSAQP 3264



Db 2905 AHSIKSDITKGEKSLPSKEVSRPVSIGIKDEKAESRRRESVAESVKPESKDATSAPP 2964  
QY 1602 -BHHPHIPISPYEHL-----LRGVSVDLYRSHIPL-----AFDPTSIPIRGIDLAAA 1647  
Db 2965 SKEHSRPSVLGSLDEGDKTTSRRVSADSTIKDEKSLLVQSEAFSEASESLK-DAEA 3023  
QY 1648 AYLPHRLAPNTYPHLYPPYLRIGYPTDALENQOTIINDYITSOQHHTATAMAQRA 1707  
Db 3024 -----PSOETSRESVTESSVKDGKSPVASKASRPASVAENAKDSADESKEORP 3072  
QY 1708 DMLRGLSPRESSLALNAAAGPRGIID-----LSQVPHLPVLVPPPTCTPAT 1753  
Db 3073 ESL-----FQSRAGSIKDEKSLASDEAKESRRRESVAEQFP-----LVSKEYSRSPAS 3124  
QY 1754 AMDRL---AYLPTAPQPFSSRRSSPLSPGGPHTLTKTTTTSSRRDRDRDRDRDR 1810  
Db 3125 VAESVKDEAKESKEESPLMSKEASRPASVAG-----SVKDEAKESKEESR-RESVA 3174  
QY 1811 EKSILTTTVEHAPIWRPGTQSSGSGSGGGGSSRRPASHAHQHSPISPRTQDA 1870  
Db 3175 EKSPLPSKEASRPASVAESVKDEADKESRRRESGAESKSLASKEASRPASVAESIKDE 3234  
QY 1871 LQ--ORPVSVLHNTGKGIITAVEPSKPTVL-----RSTSTSPVRPAATFPPTHCH 1919  
Db 3235 AEKESRRRESVAESKSLPSKEASRPTSVAKSVKDEAKESKEESRDSVAEKSP----- 3289  
QY 1920 PLGGTLGVIPTLMEPIVLLPEAPVARPERPRADTGHAFIAKPPARSGLPASP--S 1976  
Db 3290 -----LASKEASRPASVAESVQD--EAESKEESRRRESVAESPLAYK 3330  
QY 1977 KGSEPRPLVPVSGHATARTPAKNLAPHASDPAPPASADPHREKTQSKPFSIOEL 2036  
Db 3331 EASRPASVAESIKDEAKESKEESRR-----ESVAESKSLASKEASRPTSVAES 3378  
QY 2037 ELRSIGYHCSSVSPGVPVSPVSPSLTHDKGLPKHL-EELDKSHLEG---ELRPKQPG 2092  
Db 3379 VKDEAKESKEESRRDSVAESKSLASKEASRPASVAESVQDEAKESKEESRRRESVAESK 3438  
QY 2093 PVKLGGEAAHLPHL-----RPLPESOPSSPLLTQAPGVKGHORVVTLAQHISEV 2142  
Db 3439 ASKEASRPASVAESVKDDAKESKEESRRRESVAESKSLAS-----KEASRPASVAESVKDE 3493  
QY 2143 I---TODYTRHHQQULSAPLAPLYSPFGASCVLDLRPPSDLYLPPDPHGAPARGSPHS 2200  
Db 3494 AEKESRRRESVAESKSLPSK-----EASRPTS-----VAESVKD 3529  
QY 2201 EGGKRSPEPNKTSVLGGGEDGTEPSPGTEPGHRSANVYPLLYRDEQTEPGRMSGK 2260  
Db 3530 EAESKEESRRRESVAESKSLASKEASRPASVAE-----SVKDEAKESKEESRRRESVAEK 3583  
QY 2261 SPGNTSQPPAFPSKLTESNAMSAMVKKQEIKNKLNTHNRNE-----PEYNISOPGTEI 2313  
Db 3584 SP-----LASKASRPASVAESVKDEAKESKEESRRRESVAESKSLPSKESARP----- 3631  
QY 2314 FNMPAITGTLMTYRSQAVQEHASTNMGLEATIRKALMGKYDQWESPLSANAFNPLNA 2373  
Db 3632 -----TSAVESVKDEADKSK-----EESRRRESGAESKSL-- 3660  
QY 2374 SASLPAAMPITAAADGRSDHTLTPSGGGKAKVSGRPSSKAKSPAGLASGDRPPSVS-S 2432  
Db 3661 -ASMEASRPTSVAESVKDET-----EKSKEESRRRESVTEKSLPS-KEASRPTSVAES 3711  
QY 2433 VHSEGDCHNR-----RTPLTNRVWEDRPSSAGST 2460  
Db 3712 VKDEAKESKEESRRRESVAESKSLASKE-SSRPASVAES 3748

RESULT 6  
ID QY556 PRELIMINARY; PRG: 3261 AA.  
AC QY556;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 357.0 KDA PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rhodes S., Huckle E.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AL096858; CAB51072.1;  
DR InterPro; IPR000504;  
DR Pfam; PF00076; rim; 2.  
DR SMART; SM00360; RRM; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 3261 AA; 357039 MW; F8BB6A645DD9B6BC CRC64;  
  
Query Match 4.3%; Score 569.5; DB 4; Length 3261;  
Best Local Similarity 19.6%; Pred. No. 2.8e-20;  
Matches 568; Conservative 349; Mismatches 973; Indels 1011; Gaps 137;  
  
QY 71 EQPQNGERQOELHLRPE-----SHSYLPELGKS-----EMEFIESKR---PRLLELP 114  
Db 739 ERKSGQESHVNTTEKIGIDIDHTQSYRKQMEQSRKQOQMEMETIAKSEKFGSPKDV-- 796  
QY 115 DPLLSPSLATGQAGSSEDLTKDRSLTGKLEPVSPPSPHTDPELELVPPRLSKEELIQ 174  
Db 797 DEYERSLVHEVGK--PDQVTD-----SPSKK-----K 825  
QY 175 NMDRVDEI-TWVEQOISKLKKQOOLEEAAKPP-----EPEKVPSPPIES----- 221  
Db 826 RMDHVDICTKERNYRSRQISEDSERTGGSPSVRHGSFHEDEDPGSPRLLSYKGS 885  
QY 222 -----KRSVLQIYDENRK--AEAHRILEGLGPOVELPLYNQPSDTRQ 265  
Db 886 KYDEKVLPSYNTVREESLKFNPYDSSRREQMADMAKIKLSVLNSDELENRW---DSQM 941  
QY 266 YHENTKINQAMKKLTLFYKRRNHARKQWKQFCQRYDOLMEALEKKVKRIENNNRR-- 322  
Db 942 KODAGRFDSFNSII---KDSLKRKSVRL--EPGEVPSDSDEGEKSHSPASAL 995  
QY 323 -----RAKESKVREYEQKPEIRKQRELQERMQSRVQSGSL-----SMSAAR 367  
Db 996 YESSRLSFLLRDREKLRERDERSLSSLRNKFYSFALOKTTPDTKALLERAKSLSSR 1055  
QY 368 SHEVSEIIDGLSE---OENLEKQMRQAVIPPMYLDADQQRKIFINNGLMADPMKYI 423  
Db 1056 EEN--WSFLDWDSDRFANFRNNKDKKVDSDAPRIPSWYMKKKIR-TDSEGMDDKEDH 1112  
QY 424 KDOVNMNMWSEQKETFREKFMQHPKNFGLIASFLERKTVAECLVLYLTKKNENTKSLV 483  
Db 1113 KEE-----QRELQELFASRFL-HSIFQDSKRLQ-----HLERKEEDSDFTS 1154  
QY 484 RRSYRRRKSGSQOQQOQQOQQOQQOQQOQ-----MPSRQEKDEKEKEKEAEKEE 533  
Db 1155 GRIY--GK--QTSEGANSTTDSIQEPVVLPHSRPMELTRMQQKEKEDQKPEVEKQE 1208  
QY 534 EKPEVENDKEDLLKKTDDTSGEDNDEKAVASKGRKTANSQGRKGRITRSMANEA 593  
Db 1209 ---DTN-----HPKTPESAPENKD-----SELKTP 1231  
QY 594 EAITPQOQSAELASMELNESSRWTEEMETAKKGLLEHGRNWSAIRMVSGKTVSQCKNFV 653  
Db 1232 PSVGPPSV-----TVVTLESAP-----SALEKTTGDKTV----- 1260  
QY 654 FNYKKRQNLDEILQOHLKMEKERNARKKKKAPAAASEEAFPPPYVEDEEEMASVSGN 713  
Db 1261 -----EAPLVTTEKTVETPATVSEAKPASEPAPA 1289  
QY 714 EEMVEEAEALHASGNVEVPRGCSGPATVNNSSDTSIFSPHTEAAKTGQNGPKPATL 773  
Db 1290 PVEQLEQV-----DLPPGA-----DPDKAAMMP---AGVEEGSGDQPP-YL 1328

QY 774 GADGPPGPTTPTTPTSRAP--IEPTPASEATGAPTTPAPPSPAPPVVPKKEKEET 831  
Db 1329 DAK-----PPTGASFQAESNVDEPDS-----TQPLSKPAQKSEANEPEAKPDAT 1377  
QY 832 AAAPP-----VEEGEQKPPAAEELAVD---TGKABEPVKSECTEEAEGPAKGDADAA 883  
Db 1378 ADAEPDANKAEAPESQPPADELEVDPVAAKDKKPNKSKRSTPVQAAAVSIVEKPV 1437  
QY 884 EATAEGALKAEEKGGSPATTAKSSGAPQSDS--SATCSADEVDEAEGDKNRLLSRP 942  
Db 1438 TRKSEKIDREKLKRSNPRGEAKLLEKMEAEKTRTASKNSAADLE-----HPEP 1489  
QY 943 SL-LTPT-----GDPRANASPOK-PLD-----LKLKORAAAIPPIQVTK 980  
Db 1490 SLPLSRTRRNVSYATMGD--HENRSPKPEVEOPRVTRKRLERELQEAAYV--TTPR 1546  
QY 981 VHEPPR-----EDAAPTKPAAPP-----PQNL-----QPESDAP 1012  
Db 1547 RGRPKTRRRADEEENEAKPEATLKPPEGWRSPRSQKTAAGGGPQGGKKNPKVDAT 1606  
QY 1013 QPGSSPRG-----KSRSPAPPADKEAFAAEAKLPGDPCHWTSGLPVPPPREVIKASP 1067  
Db 1607 RPEATEVGPQIGVKESSMEPKAAEEAGSEQDRKD--AGTDKNPPTAPAVEVEKPK 1664  
QY 1068 HAPDPSAFSYPAGPHPLPLGLHDTARVLPPTTISNPPPLISSAKHPSVLERQIGAISQ 1127  
Db 1665 -APERKNS-----KRGSRNSRLAVDKSASLKN--VDAVSPRGAAQAAGERES 1711  
QY 1128 GMSVOLHVPSYSEHAKAPGVPMGLPLMPD-----KKLAPFSG-----VKOEQL-- 1172  
Db 1712 GV-VAVSPKESPOKEDGLSSQLKSDVPDVKPEKEDVSASGSPEATOLAKQMELEQ 1770  
QY 1173 -----SPRGAAGPESLGVTAQEAASVLRGTALGSV----- 1203  
Db 1771 AVEHTAKLAESASAAKADAPDEGLA--PEDRDKFAHQASETELAAGIINDISGPE 1828  
QY 1204 -----PGGSITKIGIPSTRV-----PDSAITYRGSIHTGTPADVLVKGTI--TRIIGED 1250  
Db 1829 NFAPPYPGESQTDLQPPAGAALQPS-----EGMETDEAVSGILETEAAYES 1878  
QY 1251 S-----PSRLORGREDSPKGVHVIYEGKKGHVLSYEGGMSVTCQSKEDGRSS 1298  
Db 1879 SRPPVNAPOPSAGPTDKTEARGNSSETSHSVPKAGSK-----EVEVTLVRKDKGRQ-- 1930  
QY 1299 GPPHETAPKRYDMWEGRVGRAISSASIEGLMGRAPRHSRPHLKEQHHRG---SI 1355  
Db 1931 -----KTRSRKRNTNKKV-----VAPVESHVP-----ESNOAQGESPA 1966  
QY 1356 TQIGIPRSYEAQEDYLRREAKLLKREGTPPPPPSRDLTEAYKTQALGPLKPAHEGLV 1415  
Db 1967 NEGTTVQHPEAPQ-----EEKQSEKPHSTPPQSCSTSLKIPSTE----- 2006  
QY 1416 ATVKBAAGSIHIEPRELRHT-----PELPLAPRLKEGSITQGTPLKYDTGASTGSKK 1470  
Db 2007 -----NSSQBIISVEERTPTKASVPDPLPPPQP-----APVDEEPOA-----R 2044  
QY 1471 HDVRSILGSPGTFP--PVHPLDVNADARALACRYESLSKRPCTASSSGGSIGARGAPV 1528  
Db 2045 FRVHSIIESDPVTPSPDSIPIPTLPSV-----TAAKLSPPVAVSG--- 2084  
QY 1529 IYPELGKPRQSP-----LTYEDHGAPFAGHLRGSPVMTMREPTPRLOEGSLSS---SK 1578  
Db 2085 -----GIPHQSPPTKVTEWITRQE-----EPRAQSTSPALPPDTKASDVDT 2127  
QY 1579 ASQDRKLTSTREIAKS-----PHSTVPEHHPHIPISYEHLLRGVSGVDLYR 1625  
Db 2128 SSTLRKILMDPKYVSATSVTSTSVTTAIEAPVSAAPCLHEAPPPVD-----SKKPLEE 2181  
QY 1626 SHIPLAFDPTSIPIGIPLDAAAAYLPHRLAPNPTYPHYLPYLRGYPTDAALENRQTI 1685  
Db 2182 KTAPPVTNNSEIQASEVLVAADKEKVAIPAKIT-----SVISRMPVSDILENSQKI 2234

QY 1686 INDYITSOQHHTTATAMAQRADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVP 1745  
Db 2235 -----TLAKPAQTITGL-----VSALTGLVNVSLVP--VNALKG 2267  
QY 1746 PTPGPTATAMDRLAYLPTAPOQFSRRHSSSPSLSPGGPHLTKTPTTTSSEERDRDRD 1805  
Db 2268 PVKGSVTTLKSLVS-----TPAGPVNLKGPV----- 2294  
QY 1806 RDREREKSLISTTIVEHAPI-----WRPGTQSSGS-----SGSS 1841  
Db 2295 -----NVLTGPNVLTTPVNAATVGTVNAAGFTVNAASAVNATASAVTVTAGAVTAAS 2347  
QY 1842 GGGGS-----SSRPASHSHAHOSPTSPTQDALOQRPSPVSLHNTKMGIT 1888  
Db 2348 GGVATTGTVTMAGAVIAPSKCKORASANENSRPHGSMPIVDDRPA---DAG--SGAGL 2403  
QY 1889 AVEPSKPTVLRSTS---TSSPVRRPAA---TFPPATHCPLGGTLDGVYPTLMEPVLLPKEA 1942  
Db 2404 RVNTSEGVLLSYSQKTEGPORISAKISQIPAS-----AMDTFQGSVSKSQVKPDS 2457  
QY 1943 PRVARP--ERPADTGHAFLA-----KPPARSGLPEASSPSKSGSEPRPL--- 1984  
Db 2458 VTASQPPSKGPOAPAGYANVATHSTLVTAOTYNASPVISVVK--ADRPSL--EKPEPHLS 2515  
QY 1985 -----VPPVSGHATIAITPA----- 1999  
Db 2516 VSTPVTQGTGVKVLVQGINTPVLVHNLVLTPTSIVTTNKKLADPVLTKIETKVLPANL 2575  
QY 2000 -KNLAPHASDPDPAPPASADP--HREKTOSKPF--SIQELELSLGVHGSYSPEGVE 2054  
Db 2576 GSTLTTPHH---PPALPSKLPTEVNHVPSPGSPADRTVSHLAAAKLDAHSRPSGPGPS 2631  
QY 2055 PVSVPSSSLTHDKGL-----PKH--LEELDOKSHL- 2082  
Db 2632 SFPRASHSPSTASTALSTNATVNLAAIGIPVPOF ISSIHPEOSVIMPPHSITQTVSLSHLS 2691  
QY 2083 EGELPKPGQ-----PVKLGGEAAHLPLRPLPESQSPSSPLLQATAPGVKGHQRVVTLAQH 2138  
Db 2692 QGEVRMNTPLPSIYISIRPEALHSR-APL---QP-----QOIEVRA-- 2730  
QY 2139 ISEVITQDTRHHPOOLSAPLAPLYSPFGASCPLDLRRRPSD---LYLPPDHGAPAR 2195  
Db 2731 -----PORASTPQAP-----AGVPALASQHPPEEVHVLPLVARATAPVQ 2771  
QY 2196 GS-----PH-----SEGKRSPENKTSVLGGEDG 2221  
Db 2772 SEVLVMOSEYRLHPTVTPROVRIMVPHVAVSEQPRADGVVVPKASKAP-----QOPG 2827  
QY 2222 IEPVSPPEGMTEPGHSRAV-----YPLLYRDCQEOTEPSRMGSKSPCNTSQ-----PP 2269  
Db 2828 KEAAKTPDAAKAAPTTPAPVPVPLPAPAPAPHGE---ARILVTPSNQLOGLPLTTP 2883  
QY 2270 AFFSKLTESNMAVKSQKQEKINKKLNTNHRNE--PEYNIQOPGTEIFNMPAITGGLMTY 2327  
Db 2884 -----VVVTHGVQI-----VHSSGELFOEYRGD-----IRTY 2911  
QY 2328 RSQAVQEH---ASTNMGLEALIRKALMKYDOWEE--SPPLSANAFNPLNASASIPAAMP 2382  
Db 2912 HPPAQLTHQTQPPAASVGLSPRTKTAAGQPPPEGEPLPQPVQSTQPAQAPPCPPSQ-- 2970  
QY 2383 ITAADGRSDHTITSPGGGKAK---VSGRPSRRAKSPAPGLASG--DRPP-----SVSSV 2433  
Db 2971 -----LGQGGQPPSSKMPQVSOEAKGTQGVQEPRLPAGPANRPEPHPTQVORA 3019  
QY 2434 HSE-GDCNRRTPLNVRNWRDRPSSAGSTFPFYNPLIMKLRQAGVMASSPPPLGAGSGLA 2492  
Db 3020 QAETGPTSFPSVPSVSMKPDLPVSLPTQAPKQPLFVPTTSQ-----PSTPPG---LV 3069  
QY 2493 GPHHAWDEEPK---PLLCQ 2509  
Db 3070 LPHTFQAPAKQDSSPHLTSQ 3090



```
RESULT 7
Q9NHX6 PRELIMINARY; PRT: 1966 AA.
AC Q9NHX6;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE GRUNGE.
GN GUG OR CG6964.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyridoides; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Erker A., Roure A., Core N., Angelats C., Vola C., Fasano L.,
RA Kerridge S.;
RT "Grunge is required for proximal and ventral leg development in
RT Drosophila."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF217844; AAF34752.1;
DR FlyBase; FBgn020427; GUG.
SQ SEQUENCE 1966 AA; 208033 MW; 96AF90E2082E770C CRC64;

Query Match 4.2%; Score 552; DB 5; Length 1966;
Best Local Similarity 20.0%; Pred. No. 1.2e-19;
Matches 490; Conservative 235; Mismatches 834; Indels 888; Gaps 111;

QY 328 KYREYEK-----QPFPIKQRELOQMOSR-----VGQSGLSMSAAKSEHEVSE 374
DB :
DB 16 QVNDVYAKLPDYNISFFP-IDKETDERELERSRWSGVVADGDLMLFLRAARSAAFQG 74
QY 375 IDGLSEQENL-----EKOMROLAVIPMLYDADQORIKFINMGLMADPMKVIKDYQM 429
DB :
DB 75 MCGGLEDCCLASRDDTFINALDVLDHSDGDPGKA-----LQALVKCPVSKGIDKK-- 126
QY 430 NMWSEQKETFRKFMQHPKNEGLI-ASFLEKRTVAECVLYLTTKKENYKSLVRSYR 488
DB :
DB 127 --WTEGTRKFKIGLRQFCKNEFFRIHKDLLPHKDTPELVEFFVYLYLWKKTPCANN--NRPHR 182
QY 489 RRGKSQQQQQQQQQQQQQQQQQQQPMRPSROBEKEKEKEKEKEKEKEKEKEKEKEKEKE 548
DB :
DB 183 RRRQSALRRNRYTRANNNSNTP-----PKKEDTPEPOT----- 216
QY 549 KTDDTSGEDNDEKEAVASKGRKTANSQGRKGRITRSMANEANSEAIPTQQSALASME 608
DB :
DB 217 -----ATTATAATAASE-----TASRSPAVSKE----- 241
QY 609 LNESSRWTEEE-METAKKGLLEHGRNWSAIARMVSGKTVYSOCKNFYFNKQKRONLDEILQ 667
DB :
DB 242 --ENSLTDEDDASECDSDSLTHKRDESP-SRM-----RTRN----- 275
QY 668 QHKLMEKERNARRKKKAPAAASEAAPPVVVEDEMEASGVSGNEEMVEAEALHAS 727
DB :
DB 276 -----KOONNSSTSGNNTAGNGGNATSISSGSTGGAGGNSSSKSDQSANAV-AN 327
QY 728 GNEVPRG----ECSGPATVNNSSDTEIPSPHTEAAKDTGQNGKPKPATLIGADPPPGPP 783
DB :
DB 328 GRPKRGSTPDVSGASVDS-----PKPTTKAVAESANKRK-----GGKQDTP 372
QY 784 TPRRTSRAPTEPTPASEATGAPTPPPAPP-----SAPPPVVPK-EEKEEETAAAP 835
DB :
DB 373 NKKKRTQESNEPS-AHEENAIKEKKRPDSPVESMNSDRPSVLDGDSNTTDTTA- 430
QY 836 PVEEGEQPPAAELAVDTGAEVPKSECTEEAEAGPAKGDABA-AEATAEGALKAE 894
DB :
DB 431 -----EQSTKDSKE-TVSKCEEREMVTNDLEAKAE--KAIKALAEADSKDSAIGNM 481
QY 895 KKEGGGRATTAKSGAPODSS-----ATCSA--DEVD 927
DB :
DB 482 DEE-----TNIQAPSSADTSLVDGNPNALPSFVAAPITMKVPTTATVEALNASVD 532
```

QY 1847 ----SSRRPASHSHAH-----QHSPISTQDALQORPSVLHNTGKMGITAV 1890  
DB 1342 MPNSSAAAAAARAAASPASSVSLRSQSLHPVQPSLSHPS---SSALSAAAAAV 1398  
QY 1891 -EPSKPTVLRLSTSSVPRRAATFPATHCPL-GGTLGGVYP-----TLMEPVLLP 1939  
DB 1399 AERDRHALMRQOS-----PHMTTPPVSNASLMSPLSKMYAPQPGQGLGTSPPPHLRP 1452  
QY 1940 KEAPRVAR-PERPRADTGHAFKAPARSGLPEPASPSKSPRPLVPVPSGHATTARP 1998  
DB 1453 GASPPVIRHPQML-----PLPIAPGGGIPQIGVHP 1484  
QY 1999 AKNLAPHASPOPPAPPASASPHREKTSQKPSIOLELRLSLGYHGSYSPEGVPSVP 2058  
DB 1485 GQSPYH-----PLHPSVYSPHH-----PPN-----SPGYAPYGP 1518  
QY 2059 VSSPSLTHDKGLPKHLELKDLSHLEGELRPQGPVKLG-FAAHLPHLRPLP-----ES 2112  
DB 1519 -----GFPAYMKP-----PPQPGQLDPAAVMAAHAGLQGPQGMQD 1557  
QY 2113 QPSSSPLLQATPGVKGHORVTLAHI SEVITQDY-----TRHHPQLSAPLPAP 2162  
DB 1558 EQNAAAAAQAARAAQOHAQAAAAAQAQOQKAPQOQPGMPPNKPPTKTPQCGGMP-- 1615  
QY 2163 LYSFGASCPLDLRRPPSDL-----YLPDPDHG---APARGSPHS-----EGGR 2205  
DB 1616 ----PCMGPGTPTGLPGAYPCSHMPGYPPQPGHSPAPQDQPHGLKPSHMDALRA 1671  
QY 2206 SPEPNKTSVLGGEDGIEPV-----SP-----TEPGHSRA 2240  
DB 1672 HAHSAANSAGMGHGHTEPLDIEPDPEIPSPTHNIPRGPSPEAKPDDTECHRSQA 1731  
QY 2241 VYPLLRYDEQTEPSRMGSKSPGNTSQPAFTSKLTESAMVSKKQKQKLNTHNRN 2300  
DB 1732 IFVRHIDRGDYNSTR-----TDLIFKPVADSKLA---RKREDRKLAEKERE 1777  
QY 2301 EPEYNISQGTIFNPAITGTGLTYRQAQVQEHASTNGLLEAIRKALMGKYDQWES 2360  
DB 1778 RROOQOQO-----QOQOQOQOAAAAQAQAKKA-----ELK 1811  
QY 2361 PPLSAN-----AFNPL-----NASASL-----PAAMPI 2383  
DB 1812 PPVADPPALRQLSEYARPHVAFRELEIKNAQAAAAAQSRLDPHMYEYRRGHPHSPFL 1871  
QY 2384 TAADGRSDHTLSPGGGKAKYSGRESSRKAAPGLASGDRPPSV 2430  
DB 1872 YANPAISQMERERLGIPIPHVGLDPEGHVVRMPPEAGFOLPPNV 1918

RESULT 8  
Q9P3J0 PRELIMINARY; PRT; 2649 AA.  
AC Q9P3J0  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CONSERVED HYPOTHETICAL PROTEIN.  
GN B7F21.40.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Algn V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL369901; CAB97476.1; -;  
DR InterPro; IPR001005; -;

DR PROSITE; PS50090; MYB 3; 1.  
SQ SEQUENCE 2649 AA; 294398 MW; 36FB50BF3917F2AB CRC64;  
Query Match 4.2%; Score 552; DB 3; Length 2649;  
Best Local Similarity 19.3%; Pred. No. 1.7e-19;  
Matches 406; Conservative 268; Mismatches 771; Indels 658; Gaps 92;  
QY 60 QQRRRRSLLSFPQG--NERSOELHLRSHESLYLDELKSEM--EFIESKRRLLELLPD 115  
DB 848 KPVQKPAQKLDELLEPPRVSESEFVAPKPEPKVPSEPKKAPKAPKPEKLPPEQIPR 907  
QY 116 PLLRPS-LLATGQAGSEDLTKDRLSLTGKLEPVSPSPPTDPE 159  
DB 908 POPQPOQPOPOQAQAQAPPOLITEAPEKSVTEPEREAEVEQEAETQEQAP--TERQ 966  
QY 160 LELVP-----PRLSKEEL-----IONMDRVDREITMYEQOISK 193  
DB 967 LLAIPVTEENEETSFRLPEPQPKMEEMDTGGSLPLPTVVEHKVVDQDVMVD----- 1020  
QY 194 KKKQOOLEE---AAKPPPEKPVSPPIESKHSRLSVQIYYDENRKAQAAH----- 242  
DB 1021 -----VDDEGRDALQP--PCKPVLRLDYSGDGESVQFPFPNQPTSDGASSRMD 1072  
QY 243 -----RILEGLGQVPELPLYNQPSDTRYHENIKINQAMRKK--LILY 283  
DB 1073 EERTDDASVSGSVHRELSTPTTDELFPVN---VRPMHOSSKVRKLSSEOSPGFAG 1128  
QY 284 F-----KRNHARKQWKFCQRYDQLMFALEKKVERIENNRRAKESKVR 331  
DB 1129 FMGNIRDMAETAKEQHEA---KQVRYNYDAYLRF-----TLDSDP--TAVKSR--- 1173  
QY 332 YIEKOPPETRKQRELQERMQS---RVQGGSGLSMAARSEHEVSEII--DGLSE--QENLE 386  
DB 1174 ---NOPSQDKDKVSGKGHSGSDNHGKGGRRTTSRFTSLDVEYAIQESIREAQEKKE 1230  
QY 387 KOMR-----OLAVIPPMY--DADQRIKFINNGLMADPMKYVKDQVYNNM-- 432  
DB 1231 REERAQREKRYTDKEAVIPENITWDEKHQLFVDSGLL--PLE-----KLVTWQAVP 1283  
QY 433 -----SEQKETFREKPMQHPKNFLIASLERKTVAECLVLYLTKKNNYKSLVRSY 487  
DB 1284 YHVNTEAEAEKFAKAYLEYPKWGKIAHELPNRDPHSVIQIYVYAKKRELNLKRLKKOP 1343  
QY 488 RRRGKSQOQOQ-----QOQOQOQOQOQPMPPSSDEKDEKEKEKEKEEKEPEV 538  
DB 1344 RRRKKRGKQKYNALVSELGNPENETDQNGENSGNRQOQRRRAAAPSNGHEATPNA 1403  
QY 539 END-----KEDLLKKTDDTSGEDNDEKEAVASKGRKTANSQGRKGR----- 582  
DB 1404 DSDGATPATPGRRRAGTTTEPKNDSGAE---KPEGVKGGRRARQPKADKPKVPPQAQ 1460  
QY 583 -----TRSMANEANSEEAITPQOSAE----- 603  
DB 1461 AIAPTTPAASGKGTNARSNSRTRVQNPEWMPKTPAELGARGVQNDPMFQVPGSMOPPLA 1520  
QY 604 -----LASMELNE----- 611  
DB 1521 PAAQPTLASPERAPPTLASTTISEVMAPSLRPEPPAPPASLPTTFEIGOSSPRTTRTQ 1580  
QY 612 ---SSRWTEEMETAKKGLLEHGRNNSAIARMVSGKTSVQCKNFYFNKK--RQNLDEILQ 667  
DB 1581 QASSYWSVSESNDFPSLLRSFGTDSAIANHMGTKTVQVNVRYLKRQKEGQEWQIAT 1640  
QY 668 QHKLMEKERNARRKKKAPAAASEEAPPPVVEDEMEASGVSGNEEWVEAEALHAS 727  
DB 1641 EADLK--KQGERRRPPPTPSAGPR-----KRYDVPSSSGHRPLAAAEAE-- 1684  
QY 728 GNEVPRGCGSPATVNS-----SDTESIPSPHTAAKD--TGONGPKPPATLGD--GPPPG 781  
DB 1685 -PQAKSEAAPANQPFSSRFQVTPIQASVLTLLVQSRSTVMPAPLPAPVPAASASAA 1743  
QY 782 PTPPRTSRAPIEPTPASEATCAPTPPPAPSPSPAPPVVPKKEKEEETAAAPVVEGE 841

Db 1744 POTPSR-----PCPAVSQTMSPVPHPL-----RQPTATFVTEREGEPI---PV-----1784  
QY 842 EQKPPAAEELAYDTCKAEPEVKSECTEAEAGCPAKGDAEAAEATAEGKAEKGGSG 901  
Db 1785 SQPPAQTOHPVRIISOKTAPVSSMPSASEAMPASA-GWIPKANRPTPLLSQOHE----1839  
QY 902 RATTAKSSCAPODSSSATCSADEYDEAGGDKNLLSPRSLLTPTGDPRA-----953  
Db 1840 -----LREYREIREGRERQI---RMEPLQPREQPRMERPAMRF 1875  
QY 954 NASPOKPL-----DLKOLKORAAAIPIQVTKVHEPPREDAAATPKPAPPAPP----1000  
Db 1876 KOEPEQPLHDDPFQNFMPQPSQWOPRAEAPMGR-----QDPPR-SMPSAQSYTPPIQA 1930  
QY 1001 -PPONLQPEDAPQO--PGSSPR-----GKSRSPAPPADKEFAAEAKLPGDPCCWTS 1051  
Db 1931 QVNRLLSESVPQRTQPOLSPAMERPPVPSTQRMPTSMQOYSAN-----TS 1979  
QY 1052 GLPFFVPPPREVIKASHAPDP-----SAFSY-----APGHPPLPLGLHDTARVLPRLPPTI 1102  
Db 1980 AAQ-PVPPQAPPAIP--PEPKVSSIFSILNDNDPPAPAPAKRVNDVAS--MPRAAST 2034  
QY 1103 SNPPPLISSAKHPSVLERQIGAISOGMSVOLHVPVSEHAKAPVGPVTMGLPLMDPKKLA 1162  
Db 2035 STPPP-----2039  
QY 1163 PFSGVKEQLSPRGAGPPESLGVPTAQASVLRGT-ALG-----SVPGGSITKGIPS---1214  
Db 2040 -----QOMSARPPQPPPT-----TAVSASQRDEALGYIYARNPPSAAQAAMPKLP 2088  
QY 1215 --TRVPSDAITY-RGSITHGTPADVLVYKGTITRIIGEDSPSLRDRGRDLSFKGHVYI- 1270  
Db 2089 YHTQSPQPPHNMVARSMSGMEQVPSAAGDNREFSRHQYQHQPGGASNSPVPHQVHH 2148  
QY 1271 -----EGKKGHV-----LSYEGGMSVTOCSKEDGRSSGPPHETAAPKRTYDMMGRVGRA 1321  
Db 2149 YSQSAQHPOHQLQPMAYPSQQOQPYATSOALAASTPPQYAAHPSSTSGREQAQSAE 2208  
QY 1322 ISSASTEGLMGRAIPERHSPHLEQHHIRGSITQIPRSVVEAQEDYLRREKLLKRE 1381  
Db 2209 EWSSTQOQAAYVGLQROOQOQOQOQOQO-----QOQOQOQOQOQOQOQO 2254  
QY 1382 GTPPP-PPPSRDLTAYKTQAL-----GP-LKLKPAHEGLVATVKEAGRSIHEIPREELR 1434  
Db 2255 GWPPSHPTPSKSSQASVPSQTAWAAGHGNVQAKPPQMG--SAMSQOQHSWQATPTQO--2310  
QY 1435 HTPPELPLAPRLKEGSITQGTPLKTYDTGASTTGSKKHVDVRSLLGSPGRTFPVPHLDVMA 1494  
Db 2311 --PHALGLREAPRQGVAFSAHEAQSPTGGSVYVSHQH--RSLDGR--SQFPP-----MP 2358  
QY 1495 DARALERACYEESLKSRRPCTASSSGSSTARGAPVIVPELKGPRQSPLTVEDH-GAPFAGH 1553  
Db 2359 DPRDRQ-----NLRGEP--VPOGQP-----VVRVNTPGPGH 2390  
QY 1554 LPRGSP--VTMREPTPLRLEGLSSKASQDKLTSTP-----REIAKSPHSTVPEHRP 1605  
Db 2391 GPGPGPGQVPCRGEP-----ADLRMQMSQARSYTPGPGVAGFEGMGPPSSSL--GYP 2444  
QY 1606 HPISP---YHLLRGVGVLDYRSHIPIAFDSTSP-----1639  
Db 2445 EQIRDAQIRDAQIREMGSRLGRHVRMDRPRELGRDPRVAAQOQEAQOQOQO 2504  
QY 1640 GIPLDAAYLYLPHILAPNPTYPHYLPVLYRGPDTAALENRQIINDYITISQOHHHT 1699  
Db 2505 GVPANAHH-----POHIQ-----VQGHPOQHOIAHQ-----OHVQOQOQHMDM 2543  
QY 1700 ATAMA-----QRADMLRGLSPRESSLALNYAAGPRGIIDLQVPHPLVLPVPTP 1748  
Db 2544 RNVAQASHQOQVYVPHQHGGMARQLRPQHOYDQOQGHGPGPANVMGADTTALATLRKGNWD 2603  
QY 1749 GPPATAMDRLAYLTPAQPFS--RHSSSPSLSPGPGTHLTTPKPTTTSSSRERDRDRDR 1806

Db 2604 GI-----FTSNKRHGQD-----GTRKHNRRERASSETNKAEDPSORROK 2642  
QY 1807 DRE 1809  
Db 2643 KKE 2645  
RESULT 9  
QVSK5 PRELIMINARY; PRT: 1963 AA.  
AC QVSK5  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE CG6964 PROTEIN.  
GN GUG OR CG6964.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=BERKELEY;  
RC MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003555; AAF50413.1; -;  
DR FlyBase; FBgn0020427; GUG.  
DR InterPro; IPR000949; -;  
DR InterPro; IPR001005; -;  
DR Pfam; PF00249; myb\_DNA-binding; 1.  
DR Pfam; PF01448; ELM2; 1.  
DR SMART; SM00395; SANT; 1.  
SQ SEQUENCE 1963 AA; 207927 MW; 8F386DE56E66CB68 CRC64;





QY 1668 -YLIRGYPDAALENRQTIINDYITISQMHNTATAMAQADMLRGLSPRESSLAL---N 1723  
Db 1211 EKLIFGLPLSPVARGSGNSVGMEDGPRRRRGRQAQDDK-----PPRFRLQKQREN 1266  
QY 1724 YAAGPRGIDLSQVHLPLVLPPTTGTATAMDRLAYLPTAPQPPSSRHSSPLSPGGPT 1783  
Db 1267 AARGSEKPSLT---LASAP---APEEA---LTTVTVAPAPRAAAKSPDLSNQSD 1315  
QY 1784 HLTKPTTSS-----SERDRDRE-----RD 1805  
Db 1316 QANEWEATASESDTTSRRGDKAPPVLLTPKAVGTPGGGGGAVPGISAMSRGDLQS 1375  
QY 1806 RDREKSLTSTTTVEHAPIRWPRTEQSSSGSGSGGGG----- 1846  
Db 1376 RAKDLSKRSFSSQRMERQNRPPGGKAGSGSGSGGGGGPGGRTGPGRGDKRSWPS 1435  
QY 1847 ---SSRRPASHAHQSPISPTQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSSTTS 1904  
Db 1436 PKNRSRPPEERPPGLPLPPPPSSSAVRLDQVIH-----SNPAGIQQAALQ 1482  
QY 1905 SPVRAATFPATHCPLGTLGVYTLMEPVLLKPEAPRVARPERPRADTGHAFLAKPP 1964  
Db 1483 LSSROGSVTAPGGH-----PR-HKPGPPQAQPGPS--PRPP 1515  
QY 1965 AR-----SGLEPASPSKSGSEPRPLVPVSG-----HATIARTPAK----- 2000  
Db 1516 TRYEQRVNSGL---SSDPHEEPGPMVRGVGTGPRDSAGVSPFPKRRERPPRKPELLO 1572  
QY 2001 ---NLAPHIAS-----PPPPAPPASAD-----PH-----REKTSKPSIQBLE--- 2037  
Db 1573 EESLPPHSSGFLGSKPGCGQAESRDTGTALPTWNLHTATSRKSYRPPSSMEPWW 1632  
QY 2038 ---LRSLGVHSGSYSPGEGV-----PVSVPSSPSLTHDKGL-----PKH-- 2073  
Db 1633 EPLSFEDVAGTEMQSQSGVDLSDGSQVSGGPGCQSRSP-----DGLKGAAGPPKRP 1688  
QY 2074 ---LEEDKSHLEGELRPKPGVKLGGEAAHLPHLPLPESQSPSSPLQATPQVKGH 2129  
Db 1689 GSSPLNAVCEGPGSEPPRRPPAPHPDGRKELPREQPLP-----PGPIGT 1735  
QY 2130 QRVVTLAHSIVITQDTRHHQQLSAPLAPLYSFGASCPLVLDLRPPSDL----- 2183  
Db 1736 ER-----SQH-----TDRTEPGPIRPS-HRPGPPVQF-GTSDKDSLDVLVGDLSLKAE 1784  
QY 2184 -----YLPPPDHGPAPAG-----SPHSEGGKRSPE-----PN 2210  
Db 1785 LTASVTEAIPVSRDWELPLSAASAREPQSKNLDGCHVPEPSSSQRLYPEVFGSAGPS 1844  
QY 2211 KTSVLGGGED-----GIEPVSP-----PEGMTEPGHSRAV-YPLLYRDG 2249  
Db 1845 SSQISGGAMDSQHPNSGGFRPGTSLPHYRSOPLYLPPGAPPALLSGVALKGQFLDF 1904  
QY 2250 EGTEPSRMGSKSPGNTSOPPAFF-----SKLTESNAMSVMKSKQBINKLNTHRNE 2301  
Db 1905 STMQATELKGPAAGLVLPFPSPFLYSFAPCFSPPLDTSLLVQV----- 1948  
QY 2302 PEYNISQSGTEFNMPATIG--TGLMTYRSOAVOEHASTNMGLEAIIRKALMGKYDOWEE 2359  
Db 1949 ---DLPSF-SDFYSTPLPLOGGSGFLPSGAPAQ-----MLLPMVDSQLPVNFGSLPP 1998  
QY 2360 SPPLSANAFNLNASLPAAMPITAAADGRSDHTLTSPGGGKAKVSGRPSRRKAKSPAP 2419  
Db 1999 APP---PAPPPLSLLPVGALOPPSLA-----VRPPAPATRVLPSPA- 2038  
QY 2420 GLASGDRPPSVSSVHSE-----GDCNRRTPLTNRWEDRPSSAGS 2459  
Db 2039 -----RPPFASLGRAELHPVELKPFQDYQKLSSNLGPGSSRTPTTGRSFGSLNRLKA 2092  
QY 2460 TPFYNPILI---MRLOAGVMASPPP-----PGLPAGSGP---LAGPHHAWDEEP 2502  
Db 2093 TFSYSGVFTQRVDLYOQASPPDALRWIPKWPWERTGLPPREGFSRRAEPEPSRGDKPEP 2151

RESULT 11  
Q9VRY3  
ID Q9VRY3 PRELIMINARY; PRT: 3080 AA.  
AC Q9VRY3;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CG10115 PROTEIN.  
GN CG10115.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Beres P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacle J.M.,  
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003561; AAF50647.1;  
DR Flybase: FBgn0035712; CG10115.  
DR InterPro: IPR001487;  
DR InterPro: IPR001487;  
DR Pfam: PF00439; bromodomain; 1.  
DR PRINTS: PR00503; bromodomain;  
DR PROSITE: PS00455; AMP\_BINDING; 1.  
DR PROSITE: PS00114; bromodomain\_2; 2.  
DR SMART: SM00297; BROMO; 1.  
SQ SEQUENCE 3080 AA; 338176 MW; 4602054387DE2C12 CRC64;

Query Match 4.0%; Score 533; DB 5; Length 3080;  
Best Local Similarity 18.6%; Pred. No. 1.8e-18;  
Matches 561; Conservative 385; Mismatches 1041; Indels 1022; Gaps 138;







QY 933 DKNRLSPRLTTPGDP--ANASPOKPLDKOLQ--RAAAIPIQVTKVHEPPPRD 988  
Db 933 SS-----SPSPSRVTSRTTPRRSRVSPCS NVESRLLPYSHSGSSSDPTKVKEPTPPQS 988  
QY 989 -----AAPTCKAP-----PAP-----pp-----1001  
Db 989 HSGSISPYKVAQTPPGPSLGSKSPQEKSDSLVQSCPSLSLACAGVKSSTPPGES 1048  
QY 1002 -----PONLOPESDAP---QOPGSSPRG-----1021  
Db 1049 YFGVSSLQKGSQSPDHRSDTSPEVQSHSESPQSLOKQTSFKGGRSSSPVTEL 1108  
QY 1022 KSRSP-----APPADKEFAAEAKLPGDPPCWTSGLP-----1054  
Db 1109 ASKSPIRODRGEFASAPMLKSGMSPQSRFQSD-----SSSYPTVDSNLSLQOSRLETAES 1164  
QY 1055 ---FVPPREVTKASPHADPSAFSYPGHPPLGLCHDHTARVPLPRPPTINPPPLISS 1111  
Db 1165 KEKMALPQEDATASP-PRQDKKFS-----PPVQDRPESSLVFKDTRLTPPRRSRG 1215  
QY 1112 AKHPSVLERQIGAIISQGMVOLHVPSYSEHAKAPGVPTMGLPLPMDPKKLAPFSGVKQEQ 1171  
Db 1216 AGSSPETKEQNSALTSQDELMVEVKESEEPAGQILSHLSSELKEMSTSNF-----E 1269  
QY 1172 LSPRGQAGPPELSGYPTAQEAASVLRGTALGSV-----PGGSITKGIPSTRVPSDS 1221  
Db 1270 SSPEVEERPAVSLTLDOQSQAASLEAVEVPSMASSWGGPHFSPHEKELSNPLRENSFGS 1329  
QY 1222 AITYRGSTHTGTPADVLYKGTITRIIG-----EDSPSRLDRGREDSLPKGHVYEGKK 1274  
Db 1330 PLEFRNSGPLGTEMGTFSSEYKEDLNGPFLNQLQETDPS-LDMKEQSTRSSGHSSSELSP 1388  
QY 1275 GHVLSEYEGMVSYTOC-----SKEDGSSSGPPHETAAPKRTYDMMEGRVGA 1321  
Db 1389 DAV--EKAGMSNQSISSPVLDAVPTFSRERSASSPEMKDGLPRT-----SR 1438  
QY 1322 ISSASIEGLMGRAIPPERHS-----PHHLKEQHRIHG- 1353  
Db 1439 SRSGSSPGLRDSGTSPSRHSLSGSGSPGMDIPRTFPRGRSECDSSPEPKALPQTPRPSR 1498  
QY 1354 -----SITOGIPRSYEA---QEDYLREAKLLKREGT-----PPPPPSR-- 1391  
Db 1499 SPSSPELNKCLTPQRESRSGSESSVDQVARTPLGQRSRSGSQELDVKPSASPOERSE 1558  
QY 1392 -DLTEAYKTQALGPL-----KLKP--AHEGLVATVKEAGRSIHEIPRE 1432  
Db 1559 SDSSPDSKAKTTPLRQRSRSGSSPEVDSKSLSPRRSRSGSSPEVKDKPRA---APRAQ 1615  
QY 1433 --LRHTPELPLARP-----LKEGSITQGTPLKYDTGASTTGSKKHVRSRSLIGSPG-RT 1483  
Db 1616 SGSDSPE-PKAPAPRALPRRSRSGSSKGRGSPGEGSSSTESSPEHPPKSRTARRGSR 1674  
QY 1484 FPPVHPLDMADARALACRYEESLKS---RPGTASSGGGSIARCAPVIVPELCKPQS 1539  
Db 1675 SPEPKTKSNTPPRRSRSSPELTKARLSRSGSSASSPSTRTPP-----RHRS 1727  
QY 1540 PLTYEDHGAPFACHLPR---GSPVTM-----RPTPLRQSGSLSSSKASODRKLTSPP 1589  
Db 1728 PSVSSPEAEKSRSSRRSSASSPTKTTSRGRGSPFKPGRQLQSRSSRREKTRTRR 1787  
QY 1590 REIAKSPHSTV-PEHHHPISPYEHLRLRGVGVDIYRSHIPALFD-----1633  
Db 1788 RDRSGSSQTSRRRQSRSSRVTRRRRGSG---YHSRSPARQESSRTRRRRGRSRT 1844  
QY 1634 -PTSIPRGIPLDAAAAYLPHRLANPNYPHYLPYPLINGYDPTAALNRQIINDYITS 1692  
Db 1845 PPTSRKRSRSTSPAPWKRSSRA-----SPATHRRSRSTPLISR 1885  
QY 1693 QOMHNTATAMAQADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLVLVPPTPGPA 1752  
Db 1886 RRSRSTSPVSRSSRSTSVTRRRSRGRASPVSSRR---SRSTP-----PV 1930  
QY 1753 TAMDRLAYLPTAPOFSSRHSSPLSPGPTHLTPTTSSSERERDRDRDREREK 1812

Db 1931 TRRRSRSTPTTR--RSRSTPPVTRRRSRSTPPVT-----RRSRSTSPITRRSR 1983  
QY 1813 SLTSTTTTVEHAPIWRPGTEQSGSGSGGGGSSSR--PASHSHAHOHSPISTPTQDA 1870  
Db 1984 SRTSPVTRRRSRSTSPVTRRRSRSTSPVTRRRSRSTPPAIRRRSRSTPLPKRRSR 2043  
QY 1871 LQORSVLHNTCMKGIITA---VEPSKPTV-LRSTSTSSVVRPAATPPATHCPLGCT 1924  
Db 2044 SRSPLAIRRRSRSTPRTARGRSLTRSPAIRRRSASGSSSDRSRSATPPATRNHSGSR 2103  
QY 1925 LDGV-----YPTLMEPV-----LLPKEAPR 1944  
Db 2104 TTPVALNSSRSCFSRSPMSPTPLDCRSPGMLEPLGSSRTPMSVLQAGGSMMDGPGPR 2163  
QY 1945 VARPERPRADTGA-----FLAKPPARSGLPEPASSPKGSEPRPLVPPVSG 1990  
Db 2164 IPDHORTSYPENHAOSRIALALTAISLGTARTPPPSAAGLAARMSQVPAPVPLM-----2218  
QY 1991 HATIARTPAKNLAPHASDPDPAPPASADPHREKTQSKPFSIQEELR-----SLG 2042  
Db 2219 --SLRTAPANLASR-----IPAASAAAMNLASARTPAIPTAVNLADSRTPAAAAMNLA 2271  
QY 2043 YHGSSYSPEGV---EPVSPVSSPLTHDKGLPKHLEELDKSHLEGELRQKPGPVKLGE 2099  
Db 2272 SPRTAVAPSAVNLADPRT-PAPV-----NLGARTPAALAALSLTGS 2314  
QY 2100 -----AAHLPHLRPLPESOPSSPLL-----QTAP-GYKGHQRVVTLAQHISEVITQDY 2147  
Db 2315 GTPPTAANYPSSSRTPQA-PASANLVGRSAHATAPVNIAGSRTAAALA-----2362  
QY 2148 TRHHPOOLSAPLAPLYSPGA-----SCPULD---LRRPPSDLYL 2185  
Db 2363 ---FASLTSARMAPALS--GANLTSRVPVPLSAYERVSGRTSPPLDRARSRTPPS---2412  
QY 2186 PPDHGAARGSPHSEGGKRSPEPNKTSVLGGEGDIEPVSPPEGTEPGHRSRASYPLL 2245  
Db 2413 -----APSQRMTSE---RAPSPSRM---GQAPSQSLLPP-AQDQP---RSPV---2451  
QY 2246 YRDGEQTESRMGSKSPGNTSQ--PPAFSKLTESAMVSKKOEINKKLANTHNEPE 2303  
Db 2452 -----PSAFSDQSRCLIAOTTPVAGSQSLSSGAVATTTSSAGDHNGMLSVAPGVPH 2503  
QY 2304 YNISPGTEIFNFPALTGTGLMYRSQAVQEHASTNMGLEAIIRKALMGKYDQWESSPL 2363  
Db 2504 SDVGP-----PASTG-----AQPSALAAL-----QPAK 2528  
QY 2364 SANAFNPLNASLPAAMPITAADGRSDHTLTSPPGGGKAKYSGRPSRKAAPAG---2420  
Db 2529 ERRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 2588  
QY 2421 LASGDRPP-----SVSSVHSEGDONRRTPLTNRWEDRPSSAGST 2460  
Db 2589 AVREGRPPEPTPAKRRRRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 2648  
QY 2461 PFPYPLIMRLQAGVMASPPPPG 2483  
Db 2649 PSPAKPGQALPKPASPKKPPPG 2671  
RESULT 13  
P70670  
ID P70670 PRELIMINARY; PRT; 2187 AA.  
AC P70670:  
DT 01-FEB-1997 (TReMBLrel. 02, Created)  
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,  
DE MUSCLE-SPECIFIC FORM GP220).  
GN NACA.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=96312450; PubMed=8698236;  
RX Yotov W.V., St-Arnaud R.;  
RT "Differential splicing-in of a proline-rich exon converts alphaNAC  
RT into a muscle-specific transcription factor.";  
RL Genes Dev. 10:1763-1772(1996).  
DR EMBL; U48364; AAB18734.1; -  
DR EMBL; U48363; AAB18732.1; -  
DR MGD; MGI:106095; Naca.  
DR InterPro; IPR002715; -  
DR InterPro; IPR003037; -  
DR Pfam; PF01849; NAC; 1.  
DR Pfam; PF02094; TS-N; 1.  
SQ SEQUENCE 2187 AA; 220599 MW; 003646AA864DEBFD CRC64;  
  
Query Match 3.9%; Score 520; DB 11; Length 2187;  
Best Local Similarity 19.7%; Pred. No. 5.5e-18;  
Matches 448; Conservative 219; Mismatches 782; Indels 830; Gaps 93;  
  
QY 738 GPATVNNSDTESIIPSTEAA-----KDTGQNGKPPATLGADGPP---PGPPT 784  
Db 3 GEATETVPATEQELPOQAETAVLPMSSALKVAAVQGPQPTPPSSILGPOQSPIVTAHQPS 62  
QY 785 P-PRRTSRAPIE---PTPASEA---TGAPTP-----PP-----810  
Db 63 PLPSSVSTPEFVPPAQPIATAALPSGTAPPTPFLPHLIGPPISPAALALASPMIGLA 122  
QY 811 -----APPSAPPVVPKKEKEETAAPVE-----838  
Db 123 QKGARSSAPLSLVALAPHSVQKSSVCPPHPLTSPSAGAEALGALTASIPLEPKTSTS 182  
QY 839 -----EGEEKQPP-----AAELAVDTGKAEPKVE 865  
Db 183 QVPSQGLTNLKGATPCPDVWRAFFSHLENPLASVQPLGMLSCPOTLSNTSPVKGVPISSA 242  
QY 866 CTEAEEGPAKGDAAEAETAEGALKAE-----KKEGGSGRAT 904  
Db 243 LTQSLRLNLKGPVSPPARNTAAPSILAPSTSLGCHPLHLHSSVDSPIQPPGQSGGLAV 302  
QY 905 T-----AKSSGAPQSDSSATSCADEVDEABEGDK- 934  
Db 303 SNPTSVGHGIAASCPPERCVVPALPSRLLAVIDSCAASPDDKSSAVTNELCSPPGSSNV 362  
QY 935 -NRLSPRSLTTPGDRANASPOKPLDLKQKORAAAIPIQTVKVEHP-----985  
Db 363 AGTSLSPKASLV-----PKGSNVALQPL-----VTQVPASOKTGLKEIPVSCIGATH 409  
QY 986 ----REDAPTKPAAPPAPPONLQPEDAQPOQCGSPRGKSRSPAPPADKEAFAAEAK 1041  
Db 410 HALDNPSAISVAPATHVPPT-----SSGLVSKDPAASVTSVLVPPAAHKQ 455  
QY 1042 LPGDPCWTSGLFF-PVPPREVIRKASHAPDPSAFSAPPGLPLGLL-----HDTARPVL 1096  
Db 456 FPAPPASATLGPVSPPLATEGLK---NLPIASLVNVCAPVSPAQAGLPTKRKDTTLQPLA 512  
QY 1097 P-----RPPTISNPPPLISSAKHPSVLEROIGALISQMSVOL 1133  
Db 513 PIALKESPPSSASLEVLSEDTVTKTTGGPAPVVRFA-----IAGVATTSLRA 563  
QY 1134 HVPYSE-HAKAPVGVPTMGLPL-----PMDPKLAPFSGVKQBLSPRGQAGPPE 1182  
Db 564 DSPPAVIRADSCVSNPTVSQPLKRSVTDPAAPRTAKTAPST---TSPLVPLASEGCPV 620  
QY 1183 SLGVPTAQEASVLRGTALGVPGGSITKGP-----STRVPSDAITYRGI 1229  
Db .621 ASSMALSPQNASVSETALALSP--EIPKSVFPDPPPLAEISFNARKVDVASHMESSGS 678  
QY 1230 THGTP-ADVLYKGTITRIIGDSSRLDRGDS-----LPKHVI 1269  
Db 679 ROGHPDASVTAKTGVVCL-----ADSSLDTSVSAKGSALSGASSPLYPLEVSFLPEAGLA 734

QY 1270 YEGKKGHVLSYEGGMSVTOCSKEDG--RSSGPPHETAA-----1306  
Db 735 VOGPKGSL---NKLSPTPSSKGPAPVSTGAPPSPKGPAPIVPTESSISSKQVPAEILPS 790  
QY 1307 PKRTYDMMEGRVGRAISSASIEGLM-----GRAIPPERHSPHLKQHHI---1351  
Db 791 POKTPEVTASRLISAVQSPKVDPIMSDVTPTSPKKTATAVPKDTSATLSLKSVPAVTSL 850  
QY 1352 -----RGSITQGITPR-----1361  
Db 851 SPPKAPVAPSNEATIVPTETPSLKNALAAATPKETLATISPKVTSPOKTPKSVSLKG 910  
QY 1362 -----SYVEAOED---YLREAKLLKE-----GTPPP--pP 1388  
Db 911 APAMTSKKATEIAASKDVSPSQFPKEVPLQLQHVPTSPKPSVDTLSGALTSPPPKQPP 970  
QY 1389 PSRLTEAYKTOALGPLKLK-----PAHEGLVATVKEAGRSIHEIPR---EELRHTPEL 1439  
Db 971 ATLAETPTYPKSPKPAASKKTPATPSPGVTAVPLEIPPCSKKAPKTAAPKESATSSS 1030  
QY 1440 PLAPPLKEGSITQ-----GTPLKYDTGASTTGSKKHVDVRSLLIGSPQT-----1483  
Db 1031 KRAPKTAVSKETPSKGVTVAVPLEISLPLKETSKSATPGEKS-----ASSPKRSPKTAP 1084  
QY 1484 --FPP-----VHPLDVMADARALERACEESL-----KSRPGTA---SSSGGSJAR---1524  
Db 1085 KETPPGGVTAVPELISLPKETPQNATNESLAASSQRKSPKTSVPKETPPGCVGTAMPLE 1144  
QY 1525 --GAPVIIVPELKGKPROSPLTYEDHGAPFAGHLPRGSPVPTMREPTPRLQEGSLSSSKASQD 1582  
Db 1145 IPSAPQKAPKTAVPKQIP-TPED-----AVTILAGSPLSPKK-----ASKTAAP 1187  
QY 1583 RKLSTPR-----ETAKSPHST-----VPEHHPHIPISPYEHLRGVSGV 1621  
Db 1188 KEAPATPSGVIVASGEISPSPKKTSKTAAPKENSATLPPKRSPKTAAPKETPATSEGV 1247  
QY 1622 DLYRSHIPLADPTISPRGIPLDAAAAYVLRHLAPNTYTPHLYPPYLRIGVDPDAALEN 1681  
Db 1248 TAVPSEISPS-PPTPASKGVV-ILTPKGAPNALAESPASP-----KKVPKTAAPEE 1297  
QY 1682 ROTIINDYITSOQMHNTATAMAQADMLRGLSPRESSLALNYAAGPRGIIDLSQVHLP 1741  
Db 1298 TST-----TPSQKIPKVG--PKRAS-----1317  
QY 1742 VLVPTPTGTPATMDRLAYLPT-----AQPFSSRHSSSPSLSPGPGTHLTKPTTT 1791  
Db 1318 -ATPPSKTKPTAVPKETSAPSEGVTVAVLEIPSPRKAPKTAAP-----KETPA 1366  
QY 1792 SSSRERDRDRDRDREREKSILTTTTVGHAPLWRPGTEQSSGSSGSSGGGSSSRP 1851  
Db 1367 PSEGE-----ATTAPVQIPSPRKGSKA-----GSKETP 1396  
QY 1852 ASHS-----HAQHSPISPRQDQALQORPSVLHNTGMKGIIITAVEPSKPTVLRSTSTSSP 1906  
Db 1397 TTPSEGVTAAPLEIPISSSKTSKMASPKETLVTPSSKKLSQTVGPKETSELEGATVPLE 1456  
QY 1907 VRPAATFPPIA---HCPGLGTLGVPTLMEPVLLPKAEPVARPERPRADTGHAFIAK 1962  
Db 1457 IPPSHKAPKTVDPKQVPLTSPKDAPTTLAESPSKPKKAPKTAAPPSEVTT-----V 1510  
QY 1963 PPARGGLEPASSKSGSEPRPLVPPVSGHATIARTPAKNLAPHIASPPPPAPPA--SASD 2020  
Db 1511 PP-----EKATPPKASCTASKVPV-----PAETQEVAVSSRETPVTPAVPPVKN 1556  
QY 2021 PHREKTOSKPSIQELESRLSGYHGSSYSGVEGVPVSPVSPSLTHDKGLPKHLEEDKS 2080  
Db 1557 PSSHKTSKTIELKE-----APATLPP-SPTKSPKIPSSKKAPR-----1594  
QY 2081 HLEGELRPQPGVKLGGEAAHPLHRLPES-QPSSPLLOTAPGVKGHQRVVTLAQHI 2139  
Db 1595 ----TSAPK-----EFPASPSIKPVTTLAQTPPSLOKAPS-----1627

```

QY 2140 SEVITQDTRHHQOOLSAPLAPLYS-PPGA-----SCP 2172
DB 1628 -----TTIPKENLAAPAVLPVSSKSPAAPARASASISPATAAPQAPKAPTEATIPSCK 1679
QY 2173 VLDLRPPSDLYLPDPHGAPARGSPHS-----EGGRKS-----PEPNK 2211
DB 1680 KAAATETPIETSTAPSLGAPKETSSETSVSKVLMSSPPKASSKRASTLPTATLLPSLKE 1739
QY 2212 TSVL-----GGEGDIGIPVSP--PEGWTEPGHSRSVAVPILLYRDGEQTEPSRMGSKSPG 2263
DB 1740 ASVLSPTATSGKDSHISFVSDACSTGTTTQASEKL-----PSKKG----- 1781
QY 2264 NTSQPPAPFSKLT--ESNSAMVSKQKQINKLNTNHRNEPEYNISQPGTEIFNMPAIG 2321
DB 1782 ----PTAFTTEMLAAPAPESALAITAPIQKSPGANSASSPKCPDPSSKDKTKGLPS--- 1834
QY 2322 TGLMYRSQAQVEHASTNMGLEAIIRKALMGKYDQWESPPLSANAFPLNA----- 2373
DB 1835 --AVALAPQTVPEKDTSKAETLLVSPAKGSDCLHSPKGPVGSQVATPLAFTSDKYVP 1892
QY 2374 ---SASLPAAMPITAADGRSDHTLT---SPGGGKAKVSGRPSRKAKSPAPGLASGDRP 2427
DB 1893 EAVSASV-APRPAPAA-----SLTAPSP-----VAPLPKPPILLESAPG--SVLES 1936
QY 2428 PSVSVHSEGCNRRTPLTNRWEDRPSAGSTPPYPNPLMRQAQVNASPPPPGLPA 2486
DB 1937 PSKLPVPAEED--ELPPLT-----PPEAVSGGEPFQPILVN-----MPAPKPAQTEA 1981

RESULT 14
QYVZ30
ID QYVZ30 PRELIMINARY; PRT; 3536 AA.
AC QYVZ30;
DT 01-MAR-2000 (Tremblrel. 13, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CG11122 PROTEIN.
GN CG11122.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basos P.V., Berman B.P., Bhandari D., Bolesley E.M.,
RA Beeson K.Y., Benos P.V., Berland J., Bayraktaroglu L., Beasley E.M.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler J., Brockstein P., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.D., Havelton D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin K., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

```

```

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003485; AAF48000.2; -.
DR FlyBase; FBgn0030266; CG11122.
DR InterPro; IPR000822; -.
DR InterPro; IPR002965; -.
DR Pfam; PF00096; zf-C2H2; 5.
DR PRINTS; PR01217; PRICHEXTENS.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
KW DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 3536 AA; 384634 MW; E16091990PBF9EEE CRC64;

Query Match 3.9%; Score 519; DB 5; Length 3536;
Best Local Similarity 19.0%; Pred. No. 1.1e-17;
Matches 589; Conservative 364; Mismatches 1037; Indels 1112; Gaps 136;

QY 42 YOHHSRDYASHLSPGSIIPQRRRSLSEFQP-----GNERSQ-ELHLRPESHSLVPEL 95
DB 199 FOHLSEQADKY---FACOPCLR-----FQDADHLGHQHQHSLHPEDGHATKPL 249
QY 96 -GKSE-----MEFIESKRP-----RLELLPDLRLPSPLLATGOPAGSEDLTKDRSLT 142
DB 250 MGADEPVLFRGLTQNLRLPSHRNRQKEEPPPLATPS-----SKSRSS 295
QY 143 GKLEVPSPPPHTDPELELYPRLSKELIQNMDRVREITMVEQQISKKKKQOOLSE 202
DB 296 -----RAATROQOQHQQSQSHQ 317
QY 203 EAAKPEP-EKPVSPPI-----ESKHSVLQIYDENKK---AEAHRILEGLG 249
DB 318 TAAAGNMALQESVNPDAVFRLPASMACVSSASVANSVDFDKFYKDVISNVRNQNHL 377
QY 250 PQV-----ELPLYNOPSDT-----ROYHENIKINOAMRKKLIL--- 282
DB 378 GLFTSSSTSTTSATSSASHKSDILMLHGPSQLGPGGASYPTLITADQFGTGLPLA 437
QY 283 YFKRRNHARKQKQKFCQRYDQLEALEKKYVERIENNRRAKESKVREYEQKPFPEIRK 342
DB 438 KFRRRPHTKHSWKWKW-----DYVKKFTLINE 464
QY 343 QRELQERMQ--SRVQRCG-SGLSM--SAAARSEHEVSEIIDGLSEQENLEKQ--MKQLAV 394
DB 465 GGRLVKKLKQPNYLRLDLSKLDMMTQLTMTROKHDLSL---SLTEQNEQERLLEQLNL 521
QY 395 I-----PPLYMDADQQR-IFKFINM---NGLMADP-----MKVYKDRQVMNMW--- 432
DB 522 ILDHLRLPHILEQNEQAVIKCENEEDGLSDHGDLLPOTFADESLSLQLQPROG 581
QY 433 -SEQKETFRKFMQ-----HPKNFGLI---ASFLEKRTVAE----- 465
DB 582 SSEQRQERDQRESRNOKAVLVSGEAWRPLYLICCGAKFDQKRSLEEHTFRHSI 641
QY 466 CVLYYYLTKNENYKSLVRSY---RRRGK-----SQOQOQOQOQOQO 504
DB 642 YATHYEVVGRLLAGNLLRHLPKRALGRFAAASNCINWPQIAPATVQOPKQVOPE 701
QY 505 QQ-----QQQPMPS----- 515
DB 702 EETRASSRSYSYEVSTPTPLSALEQQQPSPASGSGNASSSSSSCSFSSSASTLLSTATR 761
QY 516 -----SQEEK 520

```





Db 1258 ERENAARGADG-----KPSL-TLAASTPGPEET-----LTAATVPPPPRRRTAAKSPDLN 1307  
QY 1780 GGPThLTkTtTSS-----SERERDRRE----- 1803  
Db 1308 QNSDOANEWEETASERSSDFASERGDKETPPAALMTSKAVGTPGANAGAGPGISAMSRG 1367  
QY 1804 --RDREREKSLTSTTVEHAPITWRPGTEQSSGSSGGGGSS----- 1848  
Db 1368 DLSQRAKDLKSFSSQSRQMDRQNRPGTGKTSGGSSGGGAGPGRTGPGRGDXYR 1427  
QY 1849 -----SRPASHSHAHOSHSPISRTODALQORPSVLHNTCMKGIIITAVEPSKPTVLR 1900  
Db 1428 SWPSPKNRSRPPPEERPPGLPLPPPPSSSAVERLDQVIH-----SNPAGIOQ 1474  
QY 1901 TSTSSPVREPAATFPATHCPLGTLGDVYPTLMEVLLPKAPRVARPERP-RADTGHAF 1959  
Db 1475 ALAQSSROGNTAPGGH-----PRKPGPPQAPQSSPPPTRYD----- 1515  
QY 1960 LAKPPARSGLEPASSPSKSGSEPRPLVPVYSGHATIAARTPAKNLAPHASPDPPAPPASAS 2019  
Db 1516 ---PPRAS--SAISSDPHFEEPGWVRGVG-----TPRDSAGVNFPP--- 1554  
QY 2020 DPHREKTSKPSIQELEL---RSIGYHGSSYSGEVPEVPVSPSLTHDKGLPKHLEE 2076  
Db 1555 -RRERPPRPKPELLOETVPASHSSGFLGSKPEVPQOEESRDSG-----TEALTPHI-- 1606  
QY 2077 LDKSHLEGELREKQPCVKLGCEAAHLPLRLP-----ESQPS 2115  
Db 1607 WNLHTATSRKSYQPSIE-----PMNEPLSFEDVAGTEMSSQSGVDLSGDSQVS 1658  
QY 2116 SSPLIQ-TAP--GVKGHORVVTLAOHISEVITQDYTRHHPOOLSAPLPAPLYSFFCASC 2172  
Db 1659 SGPCSORSSPDGGLKGS-----AEGPKRPGG--PSPLNAVGESAS 1698  
QY 2173 VLDLRPPSDLYLPDPDHGAPAR-----GSPHSEGGKRSPBPNTSVLGGE 2219  
Db 1699 GSEPEPR--RRPPASHGERKELPREQPLPPGPIGTERSQRTDRGPEGP----- 1748  
QY 2220 DGIEPVSPPEGTEPG-----HSRSVAVPLLYRDE----- 2250  
Db 1749 --LRFAHRPGSOVEFGTTNKDSDCLVVGDTLKGKELVASATEAVP-ISRDWELLPSAS 1805  
QY 2251 ---QTEPSRMGS-----KSP-GNTSQPAFFSKLTESNSAMVSKKQINKKLNTNHN 2300  
Db 1806 TSAEPQKSLGSGQCVPSPSGQRPYEVFGSPGPNQV--SGAPIDSQNH----- 1858  
QY 2301 EPEYNISQPGTEIFNMPAITGTMTYRSQAOVEHASTNMGLEAIRKALMCKYQWEE- 2359  
Db 1859 -PNSGGFRPGT-----PSLHQYRSQPLYLPPGPAPPSSALLSGVALKGQFLDFGAL 1907  
QY 2360 -----SPP--LSANAFNP-----LNASASLPAAMPITAADGRSD--H 2392  
Db 1908 QATELGKLPAGGVLYPPPSFLYSAFCPSPLPDPPLLQVRQDLSP-----SDFYS 1958  
QY 2393 TLTPSGGGGKAKVGRPSRK-----AKSPAGLASGRPPSVSVHSEDCNRRRTPLN 2448  
Db 1959 TPLQPGGSGFLPSGAPQAQMLPVVDQLPVVNFGLPPA----- 1999  
QY 2449 VWEDRPSAGSTPPFYPLIMRLQAGVNASPP-----PPGLPAGS-----GPLAG 2494  
Db 2000 -----PPAPPPSLLLPVGPALQPNLAVRPPPPAPARVLPSPARPFAPSLGR 2047  
QY 2495 HHAWDEPKPLCSOYETLSDS 2516  
Db 2048 AELHPVELKPF--QDYRKLSN 2067

Search completed: September 8, 2001, 14:38:43  
Job time: 4750 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 14:38:48 ; Search time 41.42 Seconds  
(without alignments)  
7793.922 Million cell updates/sec

Title: US-09-522-753-11  
Perfect score: 12643  
Sequence: 1 MSSSGYPPNQGFSTEQSRY.....EREPAPLLSAQYETILSDSD 2440

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 segs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPREMBL16.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_unclassified.\*
- 13: sp\_vertebrate.\*
- 14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description         |
|------------|-------|-------------|--------|-----------|---------------------|
| 1          | 5395  | 42.7        | 1047   | 4 Q9NSZ0  | Q9nsz0 homo sapien  |
| 2          | 1006  | 8.0         | 3469   | 5 Q9U412  | Q9u412 drosophila   |
| 3          | 996.5 | 7.9         | 3502   | 5 Q9VXJ9  | Q9vxi9 drosophila   |
| 4          | 584.5 | 4.6         | 5327   | 5 Q76891  | Q76891 drosophila   |
| 5          | 517   | 4.1         | 3261   | 4 Q9Y556  | Q9y556 homo sapien  |
| 6          | 507   | 4.0         | 3080   | 5 Q9VRY3  | Q9vry3 drosophila   |
| 7          | 489.5 | 3.9         | 2701   | 4 Q9Y520  | Q9y520 homo sapien  |
| 8          | 487.5 | 3.9         | 2364   | 11 Q9ER21 | Q9er21 rattus norv  |
| 9          | 485   | 3.8         | 5476   | 5 Q9N117  | Q9nj17 drosophila   |
| 10         | 485   | 3.8         | 5533   | 5 Q9VPL2  | Q9vpl2 drosophila   |
| 11         | 484.5 | 3.8         | 3576   | 11 Q9Q2W2 | Q9q2w2 mus musculus |
| 12         | 484.5 | 3.8         | 5534   | 5 Q9NHN1  | Q9nhn1 drosophila   |
| 13         | 484.5 | 3.8         | 5560   | 5 Q9VPL1  | Q9vpl1 drosophila   |
| 14         | 482   | 3.8         | 5533   | 5 Q9U6C3  | Q9u6c3 drosophila   |
| 15         | 480   | 3.8         | 3394   | 4 Q9Y6V0  | Q9y6v0 homo sapien  |
| 16         | 473   | 3.7         | 2752   | 4 Q9Q035  | Q9q035 homo sapien  |
| 17         | 467.5 | 3.7         | 2910   | 10 Q9FND5 | Q9fnd5 arabidopsis  |
| 18         | 467.5 | 3.7         | 4833   | 11 Q9QYX6 | Q9qyx6 mus musculus |
| 19         | 467.5 | 3.7         | 5038   | 11 Q9QYX7 | Q9qyx7 mus musculus |

|    |       |     |       |           |                    |
|----|-------|-----|-------|-----------|--------------------|
| 20 | 465.5 | 3.7 | 2897  | 5 Q9VID9  | Q9vid9 drosophila  |
| 21 | 463.5 | 3.7 | 3111  | 5 Q9VH10  | Q9vh10 drosophila  |
| 22 | 462   | 3.7 | 4880  | 11 Q9JLT1 | Q9jlt1 rattus norv |
| 23 | 462   | 3.7 | 5085  | 11 Q9JKS6 | Q9jks6 rattus norv |
| 24 | 460.5 | 3.6 | 2768  | 5 Q9VC00  | Q9vc00 drosophila  |
| 25 | 453   | 3.6 | 2296  | 4 Q9UHA8  | Q9uha8 homo sapien |
| 26 | 450.5 | 3.6 | 2951  | 5 Q9W3Z0  | Q9w3z0 drosophila  |
| 27 | 450.5 | 3.6 | 6815  | 5 Q9I7U4  | Q9i7u4 drosophila  |
| 28 | 449.5 | 3.6 | 4900  | 5 Q9N541  | Q9n541 caenorhabdi |
| 29 | 444   | 3.5 | 7962  | 4 Q10465  | Q10465 homo sapien |
| 30 | 442   | 3.5 | 16215 | 5 Q9NFS3  | Q9nfs3 drosophila  |
| 31 | 440   | 3.5 | 5170  | 5 Q17490  | Q17490 caenorhabdi |
| 32 | 438   | 3.5 | 6994  | 5 Q17343  | Q17343 caenorhabdi |
| 33 | 433.5 | 3.4 | 1966  | 5 Q9NHX6  | Q9nhx6 drosophila  |
| 34 | 432.5 | 3.4 | 3263  | 5 Q9I7U3  | Q9i7u3 drosophila  |
| 35 | 432   | 3.4 | 1963  | 5 Q9VSK5  | Q9vsk5 drosophila  |
| 36 | 430   | 3.4 | 3257  | 5 Q9V736  | Q9v736 drosophila  |
| 37 | 429   | 3.4 | 2649  | 3 Q9P3U0  | Q9p3u0 neurospora  |
| 38 | 429   | 3.4 | 6677  | 5 Q9N435  | Q9n435 caenorhabdi |
| 39 | 425.5 | 3.4 | 3851  | 4 Q43161  | Q43161 homo sapien |
| 40 | 425.5 | 3.4 | 3926  | 4 Q9UPA5  | Q9upa5 homo sapien |
| 41 | 424   | 3.4 | 2344  | 5 Q9N3Y8  | Q9n3y8 caenorhabdi |
| 42 | 415.5 | 3.3 | 5120  | 13 Q9PU36 | Q9pu36 gallus gall |
| 43 | 414.5 | 3.3 | 102   | 4 Q9H4R4  | Q9h4r4 homo sapien |
| 44 | 412   | 3.3 | 2829  | 13 P70039 | P70039 xenopus lae |
| 45 | 409.5 | 3.2 | 2361  | 5 Q94226  | Q94226 caenorhabdi |

#### ALIGNMENTS

RESULT 1  
ID Q9NSZ0 PRELIMINARY; PRT; 1047 AA.  
AC Q9NSZ0;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE HYPOTHETICAL 113.8 KDA PROTEIN (FRAGMENT).  
GN DKEZP434M075.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Duisterhoeft A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; ALJ37641; CAB70854.1; -;  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 1047 AA; 113817 MW; B5A1EDE938B7222A CRC64;

Query Match 42.7%; Score 5395; DB 4; Length 1047;  
Best Local Similarity 99.6%; Pred. No. 1.2e-284;  
Matches 1043; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

|    |      |                                                             |      |
|----|------|-------------------------------------------------------------|------|
| Qy | 1394 | SGQSAIKHNKSLITGTPSKSRGMPPLIIVPENIKVVERGKYEDVKAGETVRSRHTSYVS | 1453 |
| Db | 1    | SGQSAIKHNKSLITGTPSKSRGMPPLIIVPENIKVVERGKYEDVKAGETVRSRHTSYVS | 60   |
| Qy | 1454 | SGPSVLSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRGSPMMNRTSDVTTPPNKST | 1513 |
| Db | 61   | SGPSVLSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRGSPMMNRTSDVTTPPNKST | 120  |
| Qy | 1514 | NHERKSTLTPTQRESIPAKSPVPGVDPVWSHSPDPHHRGSTAGEVYWSHLPTOLDPAMP | 1573 |
| Db | 121  | NHERKSTLTPTQRESIPAKSPVPGVDPVWSHSPDPHHRGSTAGEVYWSHLPTOLDPAMP | 180  |
| Qy | 1574 | FHRALDPAAAYLFQROLSPPTGPPSYQLYAMENTROTILNDYITTSQOMQNLNRPDVAR | 1633 |









Db 1296 ASAGEATTATGATAAAGVKGKPEATAGTAAGADSRPDANDPLAKTAKRAINAE 1355  
QY 656 -----TKSEACKNFYFNYKRHNLDNLLQHKQT----- 686  
Db 1356 YNAGGNSSSSNATGASAVQVTLNGFPGYQTVVMANVKASTGGDDSGANAGGAAP 1415  
QY 687 -----SRKPREERDVSOCE-----VAS 704  
Db 1416 GSLAATNASIATSGDKIVKTTTPSRAPNSTSTTAANESSGAGVNTYGHATTAGNYLGO 1475  
QY 705 TVSAQEDEDIASNEENPEDESEVAVKPSDESPENATSRGNTPEPAVELEPT----- 756  
Db 1476 KLKAAQVEGLGAGNELH-----SDYSESKRRKRFELNSGAGGAGNATSMNTSSTSGSMNLS 1531  
QY 757 -----TETAPSTPSLAVPTKKAEDSVS----- 781  
Db 1532 SHGLKANAKGSMMAKTSMASTASVVVTTPTSGASSSLSSASSMLLISAASVNSTAAG 1591  
QY 782 -TQVNDISISAETABQMD-----VQOEHSABEGS----- 809  
Db 1592 ATSSSTATTATASAIPLLLADSGNSMVNANEITALDGKDKLASCFCVCKAEACPRTRP 1651  
QY 810 -----VCD----- 812  
Db 1652 LKKGGOQYIGDETIPAGARVCNCOCKSVRSYRPNCLPTCPNPKDRAQLRNIPSL 1711  
QY 813 -----PPPATK-----ADSV 823  
Db 1712 FELAPEVRDPLMAFEQIPPHATRCACSACLMIRKLDLPQLNLTGSGGAGSGGDDTD 1771  
QY 824 V-----EVYPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQIN 867  
Db 1772 VTSSECDEREPPGSDTASVESPENLRHKSJLTMVKQOQOQOQOQOQOQOQOQOQO 1831  
QY 868 AORPEPQS-----DN----- 877  
Db 1832 QPQPPPAPOQKQSGRGGOGTPLIITPRMSSKSGGGAOTAGDNERLLPPAAGQAP 1891  
QY 878 -----DSSATCSADEVDGEPERMPF-----MDKPSLLNPTGSI 914  
Db 1892 KKQKTSEYDSATETADEENENSP-ANRQSPKVLFGHGHGHGHGHANNVAGLPPVANNM 1950  
QY 915 LVSSPLKP-----NPLDLPOLHRAAVIPPMVSCPTPCNIPGTPVSGYALYQRHIK 965  
Db 1951 GTGGVQVPGGAAGQOVNGPISM-----RREAVNNVQDCV-----FSVIERSLK 1993  
QY 966 -----AMHESALLEQORQEQI--DLBCRSST----- 991  
Db 1994 HKGPQKGGQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2053  
QY 992 --SPCGTSKSPNREWEVLQAPHQLITNLPEG--VRLPTTRP--TRPPPLIPSSKTTVA 1045  
Db 2054 YRQDPGLKQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2113  
QY 1046 SEKPSFTMGISOGTPTGYTL-----SHNOASYTQETPKPSVGSISIGLPROQESAKSA 1100  
Db 2114 -----CAGS--NGTSDSLATSLVNSH-----MGVGIHGHPMAHASSA 2152  
QY 1101 -----TLPYIKQBEFFSPRSONSQEGILLVRAQHEGVVGTAGAIQESITRGTPTS 1151  
Db 2153 GGIGVDKATIPVVKSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 2208  
QY 1152 KISVESIPSLRGSTIQGTTPALPQTGI-----PTEALVKGSISRMPIEDSSPEKG 1200  
Db 2209 --SQSHVHPAHSHTQ-HPAIPHQSSHCQHTQLOVPEPEPQTLDLSTKPPRDCGSHPTG 2265  
QY 1201 REEAASKGHVYIEKSGHILSYDNINKNAREGTRSPRTAHEISLSKRSVESVEGNKQGM 1260  
Db 2266 AGGSSSG-----SGSG-----PSSDRHHGPPPP-----TSMKHIVRS--GMYRGDTV 2310  
QY 1261 RESPVSALE-----GLICRALPRGSPHSDLKERTV-----LSGSIHQ 1299

Db 2311 TVPSLAAPSSYLIPTRSVKTIIGGGGVVPGVLP-GVPSALYLQVPVVPVPIISG---QG 2366  
QY 1300 --TPRATTESFEDGLKYPKQIKRESPPIRAFEGAITKPKPDGITTIKEMGRSHEIP-- 1355  
Db 2367 QLPRA-----GOPPPAQ-----PP-----SGRGVAKVPPK 2392  
QY 1356 -----RODILTQESKRKTEVVQSTRPIIEGSIISQOCTPIKFN 1393  
Db 2393 LSPQOAHHLHPHSHGSPSQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2449  
QY 1394 SGQSAIKHNKSLITGPKSLSRGMP-----LEIVPNIKVVERGKYEDVAGTVRSR 1447  
Db 2450 AQOQIIVHAPATAAAPS-----LSPKFDGLVQRTTPEGVSGVGP-----GASGCK 2499  
QY 1448 HTSVVSSGSPSVLRSITLHEAPKAQLSPGIYDDTSARTPVSYQNTMSRSGSPMNRSDVTI 1507  
Db 2500 HGSITQGT-----LHMP-----HLESKRYESYKSSQSRHSPAQOQPGNOL 2544  
QY 1508 PPNKSTNHERKSTLTPTQRESIPAKSPVPGVDVPPVYVSHSPFPHHRGSTAGEVYWSHLPTQ 1567  
Db 2545 PP-----PQOSS-PQAPPQGYGVGS----- 2566  
QY 1568 LDPAMPFHRALDPAAYLAFORQLSPTPGYPSQOLYAMENTROTILNDYITSQOM----- 1623  
Db 2567 ---SPTAR-----SPFAGVVEQPVL---STRQIVMHDYITSQOMQOQ 2604  
QY 1624 ---OVNLRPDVARGL-----SPREQ-----PLGLP- 1645  
Db 2605 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2664  
QY 1646 -----YPATRGIIIDLNTMPTTILVP-----HPGGTSTPPMDRITIPGTQ 1685  
Db 2665 YSSRASPADHNSGHDALASFVDVAVQOQLPVPSPKDDKSPGPSTAPGQ-----VPGSG 2719  
QY 1686 ITFPPRPYNSMSFGPHTHLAAASAERERERERERERERERERERERERERERERERER 1745  
Db 2720 PPLGPSPL-----PPHAVV-----GVAQP 2738  
QY 1746 GRPGSHGVRSPPSV-RTQETMLQORPSVFGTNGTTSVITPLDPTAOLRIMPLPAGGPS 1804  
Db 2739 PPPTAHDOORYDLTLHHHHTLVQOQIAQOQHVSRLNVAQVDMOROM----- 2787  
QY 1805 ISOGLPASRYNTAADALAALYDAASAPQMDVSKTESKHEAA-----RLEENLRSA 1859  
Db 2788 -----DOAKVRMHQHQHVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2833  
QY 1860 VSEQOQLEQKTVLEK-----RSVQCLY-----SS 1885  
Db 2834 REREREQERERERERERERERERERERERERERERERERERERERERERERERERER 2893  
QY 1886 AFPSGKPOP-----HSSVYVSEAGDKGPPPKSRVYEEELTRGKT-----ITAANEID 1934  
Db 2894 GAGGGSPGQFLASV--PETGPRSPIDRERSYRQAHGGAPEPTPQOLSQSLID 2951  
QY 1935 VIITQIASDKDA-----RE-----RGSSDSSSSSLSSHRYETPSDAIEVI 1976  
Db 2952 AIIKHEINRSNDATAGPGRFPSPFVHAPLPGRGSGGGGTGRSS-----PANVLHPM 3006  
QY 1977 -----SPASSPAPQEKLOTYQEVVYKNAQAEENDPTQYEGP 2013  
Db 3007 YLRDLRQPLDGGAGSMITAEANNKPSGSPVINIDLQERIISAAAAVAQOQOQOQOQ 3066  
QY 2014 LHVRPQOESPQO-----QOLPPSOAECMGQVPR--HRLITLADHICQIITQDFARNOV 2068  
Db 3067 ---PPSOSSOSRSHVQGLRTPTSOAGGSAPSPQOIHFKSITFGELTDSITSDYGTN-- 3120  
QY 2069 SSQTPQ-QPPTSTQNSPALSALVTPVTRTKTSNRYSPESQAQSVHHORPGSRVSPNLVDK 2127  
Db 3121 ---PHLRPPYMAVLQETQSL-PPDRWK-QNRRMQKAEAKKH-----SQOQOQOQOQOQ 3170  
QY 2128 SRGRPGKSPERSHVSSEPIPIPPQVPPVHEKQDLSLLLSORGAEP--AEQRNARS 2185  
Db 3171 OHHAQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 3214



|    |      |                                                                 |      |
|----|------|-----------------------------------------------------------------|------|
| Qy | 1159 | -----PSLRGSIQTCTPALPQTGPTPTALYKGSISRMPIEDSDSPKGRBDAASKGHVYI     | 1212 |
| Db | 1881 | AEKSKEESRRESVAEKP--LPSKEASRPASVAEST--KDEAEKSKEESRRESVAEKSPLPS   | 1938 |
| Qy | 1213 | EKSGHILSYDNIKNARE-----GTRSPRTAHEISLKRSY--ESVBNIGIKOCM--         | 1258 |
| Db | 1939 | KEASRPASVAESIKDEAEKSKEESRRESVAEKSPLPSKEASRPASVAESIKDEAEKSKEE    | 1998 |
| Qy | 1359 | SMRESVP--SAPLEGLICRALPRGSHDLKERTVLSGIMOGTTPRATTESPEDGLKYPKQ     | 1317 |
| Db | 1999 | SRRESVAEKSPLPS-----KEAS-----RPAESVAESIKDEAEKSKE                 | 2034 |
| Qy | 1318 | IKR-----ESPPIRAFEAGAITKGPYDGIITTIKEMGRSIIHELIPQDILITQBSRKTPEVQ  | 1372 |
| Db | 2035 | ESRRESVAEKSPLPSKEAS-----RPAESVAESIKDEAEKSKEESRRESVAEKSPLPSK--E  | 2088 |
| Qy | 1373 | STRPIEIGSISOGTPTIKFONNSGQSAIKHNKSLITGPKSLRG--MPLEIVPENIKWVE     | 1431 |
| Db | 2089 | ASRP---ASVAESVKDEADKSEESRRESMAES---GKAQSIKGDQSPLEKVESRPESVAE    | 2142 |
| Qy | 1432 | RGXYEDVKAGETVRSRHTSVSS-----GPSVLRLSTLHEAPKAAQ--                 | 1470 |
| Db | 2143 | SVKDDPVKSEK--SRRESVAGSVTADSARDQSPLESKSGASRPESVVDVSKDEAEKOE      | 2200 |
| Qy | 1471 | -----LSPCIYDDTSART--PVSQYNTMSRGSPMMNRTSDVTIPPNKSTNHERRKS        | 1519 |
| Db | 2201 | RRESKTESVIPPKAKDDKPKVLPQVPSMTETI-----REDADQMPKPSQAESRRE         | 2251 |
| Qy | 1520 | TLPTQRESIPAKSPVPGVDVPSVSHPDHHRGSTAGEVYV-----                    | 1561 |
| Db | 2252 | SIA-----ESIKASSPRDEKSPIASKEASRP--GSAVESIKYIDLOKPOIKDDKSTHESR    | 2304 |
| Qy | 1562 | -----SHLPTQLDPAMPFHRALDPAAAYL-----FORQLSP-----TPGYPSQ-----      | 1600 |
| Db | 2305 | RESLEDKSAVTSEKSVSRPLSVASDHEAAVAIEDDAKSSISPDKSRPGFVAETVSSPIE     | 2364 |
| Qy | 1601 | -----YQLYAMENTROTTLINDYITSQOMQVNLRP--DVARGLSPREOPLGLPYPATRGII   | 1653 |
| Db | 2365 | EATMEFSKIEWEKSSIALSLGSGGKLOTDSPPVDAEG-----DFSHAVA               | 2412 |
| Qy | 1654 | DLTNMPTTILVP-----HPGCGTTPPMDRITYIPGQITFPFPRPNYNSASMSGHPHTLAAA   | 1709 |
| Db | 2413 | SVSTVPTTLTKPAELAQIAAKT-----VSSPDEALRTPSAPEHISRA                 | 2456 |
| Qy | 1710 | AS-----AERERERERERERERERERERERERERERERERERERERERERERERERER      | 1758 |
| Db | 2457 | DSPAECASEEIASODKSPQVLKESSRPAPAWAESKDDAAQLAKSSVEDLRSPVASTEISRPA  | 2516 |
| Qy | 1759 | PSVYRTOETMLQBPSPVF--QGTNGTSVITYPLDPTAQLRTMPLPAGGPSISQGLPASRYNTA | 1817 |
| Db | 2517 | SAGETASSPIEAPKDFAEFEQAEKAVILP-----TIELKGNLPLTSSPVDVAHASVQ       | 2569 |
| Qy | 1818 | ADALAAL--VDAASAPOMDVSKT-----KESKHEAARLEENLR--SRSAAVS            | 1861 |
| Db | 2570 | PAELSKVDIEKTASSPIDEPAPKSLTGSPAERPERPESAEKADAESVEKSKDASRPPSVV    | 2629 |
| Qy | 1862 | EOOQLE-----OKTLEYEKRSVOCLTSSAFPS-----GKQPQHSSV                  | 1898 |
| Db | 2630 | ESTKADSTKGDISPSPESVLEGPKDDVEKSSRPPSPVSASITGDTSKDVSRP---ASV      | 2686 |
| Qy | 1899 | VYSEACKDGPPPKRYEDELTRGKTITITANFIDVITITQIADSKDARERGSGQSDSS       | 1958 |
| Db | 2687 | V--ESVQDEHDKAESRRRESIAK-----VESVIDEAGKSDKSSQSDSQ--KDEK          | 2731 |
| Qy | 1959 | SSLSSHRYETPSDATE-----VIS-----PASSPAPPOEKLOTYQPEVVK              | 1998 |
| Db | 2732 | STLASKEASRRRESVVESSKDDAEKSESRESPIVASGEPVPRESKPSLDSKDTSRPGSWE    | 2791 |
| Qy | 1999 | ANOAEKNDPTROYELGHLHYRPOQESPSPOQOLPPSSQAEAGMGVPRTHRLITLADHCQI    | 2058 |
| Db | 2792 | SVTAEDEKSEQ-----QSRRESVAESVKADTKDKGKSQE-----ASRPSSVDEL          | 2835 |
| Qy | 2059 | ITODFARNQVSSQ--TPQOPPTSTF--QNSPSALYSTVPRTKTSNRYSPESQAGSVHHQR--  | 2114 |

```

RESULT 5
Q9Y556 PRELIMINARY; PRT; 3261 AA.
AC Q9Y556;
DC 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DI 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 357.0 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID=9606;
RN [..]
RP SEQUENCE FROM N.A.
RA Rhodes S., Huckie B.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL096858; CAB51072.1; -.
DR InterPro: IPR000504; -.
DR Pfam: PF000076; rim; 2
DR SMART; SM00360; RM; 1.
KW Hypothetical protein.
SQ SEQUENCE 3261 AA; 357039 MW; F8BBG6A645DD9B6BC CRC64;

Query Match 4.1%; Score 517; DB 4; Length 3261;
Best Local Similarity 20.2%; Pred. NO. 3.le-19;
Matches 546; Conservative 306; Mismatches 1038; Indels 816; Gaps

Qy 49 EVSQAQLQQOQQOQLR-----RRPSLLSEFFHGPSGRPOERRTSYEFPHPGPSP--- 98
 || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :|
Db 697 EVQSKEPIPKQLKLOVLDDQGPERDYKNYCSLRDTPERKSQEKSHSVNTEEKI 756

Qy 99 -VDHSLSKKRPLEQVSDFORVSAANLVPLHLPEGLRASADAKDPAFGGKHAPS 157
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :|
Db 757 GIDIDHTQSYRKMQEQSRRRKQMMEITA-----KSEKFQ----SPK 793

Qy 158 SPIS:-----GPCGDGNASPSKLKBELIQSDRVDRDREI-AKVQOIILKKX 205
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :|
Db 794 KDVEYERRSLIVHGVPQGVTDSDSPSKKR-----MDHVDFDICTKERNYRSSROI 848

Qy 206 QQQLLEEAANKPP-----EPKPV-SP-----PPEOK-----HRSTVOII 239
```

Db 849 SDSRGTGSPVRRHSGSFHEDEDPICSPRLLSVKGSPKVDKVLPSYNTITREESLKNP 908  
Qy 240 YDENRKK--AEEAHKIFEGGLGKVELYNOPSDKYVHENIK-----TQVNRK 287  
Db 909 YDSRREOMADMA-----KIKLSVNSEDELNRWDSQMKDAGRDFVSFPNIIKR 959  
Qy 288 KLILFFKRNRHARKQKICORYQDLMEAWKVKVDRTENNPRRKAKESKTRREYVEKOPF 347  
Db 960 DSL-----KRSVRDLEGEVPSDS--EDGEHK-----SHSPRASALYESRSLFLLRDR 1008  
Qy 348 EIRKROEOQRQVRGQA-----GLSATAIKARHEISEIIDGL--SQEN----- 392  
Db 1009 E-DKLRERDERLSSLRNKKFYSFALDKTITPDKALLERAKSLSSREENWSFLDWSR 1067  
Qy 393 -----NEKOMQLSVP---PMFDAEQRRVKFINNGLMEDPMKVVYKDRQFMNVTWTHE 444  
Db 1068 FANFRNNKDKKVDOSAPRIPISWY--MKKKIR--TDEGKMDDKKEDHKEE-----QER 1119  
Qy 445 KEIFDKFTQHPKFNGLIASYLERKSVDPCLVLYVLTCKNENYKALVRNRYGKR----- 498  
Db 1120 QELFASRELH-----SSIFEQ-----DSKRLQHLERKEEDSDFISGRYIGKQTEGAN 1167  
Qy 499 -----RGRNOIARPSQOEKVEKEEDKAEKTEKKEEKKDEBEKEDSK 545  
Db 1168 STTDSIQEPVVLFSRFMELTRMOQKEKQKPKVEKQEDTENHPKTPESAPENKDS 1227  
Qy 546 ENTKEKDKIDGTAEETEERQATPRGTANTANOGRRKGRIITRMTNEAASAAAAATE 605  
Db 1228 --LKTPPSGPPSVTVTLESAPSALERTGDKTVEAPLVTEETKEVTPATVSEEAKPASE 1285  
Qy 606 EPPPP-----LPP----- 613  
Db 1286 PAPAPVEQLEQVDLPPGADPKDEAAMMPAGVEEGSSGQPPYLDKPTPGASFQAESN 1345  
Qy 614 -PPEPISPEVETSRWTEEMEVAK-----KGLVEHGRNMAAIA----- 651  
Db 1346 VDPEPDSQPLSKPAQKSEANEKAEKPDATADAEPAQKAEAPESQPPASEDLEVD 1405  
Qy 652 -----KWGTCKSEAOCKNFYFNYKRRHNLNDLLOQHOKTSRKPREERD----- 695  
Db 1406 PPVAAKDKPKNKRSTKTPVQAQAAVSIYVEKPVTRKSERIDREKLKRSNPRGEAQKLEL 1465  
Qy 696 VSQCSVASTVSAODEDIE----- 715  
Db 1466 KMAEKITRTASKNSAADLEHPEPSLPLSRTRRRNRVSVIATMGDHNENRSPVKEPVEQPR 1525  
Qy 716 -----ASNEEENPEDESEVAEYKPSD--SPENAT 742  
Db 1526 VTRKRLERELQAAAVPTPRGRPPKTRRRADDEEENEKAEPAETLKPPEGWRSPRSQK 1585  
Qy 743 S-----RGNTPEAVEL---EPTTAPSTSPSLAVPSTKPAEDSVETQVNDLSIA 790  
Db 1586 TAAGGPGOGKKGKNEPKVDATPEATTEVGQIGVKESMEPKAAEEAGSEQKDRKDA 1645  
Qy 791 ETAQOMVDQOEHSAEBSGVCDPPATKADSVDEVRVVENHASKVEGDNTKER----- 844  
Db 1646 GTDKN-----PPETAPVEVWEKKPAKPKNSKRSKRSNRLAVDKS 1687  
Qy 845 -DLDRASEKVEPRD-----EDLVAAQOINAOPE--POSNDSSATCSADE-DVDEP 893  
Db 1688 ASLKNVDAVSPRGAQAQAGERSGVA--VSPKESPEQKEDGLSOLKSDPDVDPDQEP 1745  
Qy 894 ERORFPMDSKPSL-----LNPTGSILVSSP--LKNPNLDLPLQ 930  
Db 1746 EKEDVSASGSPSEATQAKOMELEQAVEHIAKLAESAASAAKADAEGLAPEDRDRPAH 1805  
Qy 931 Q-----HRAAVTPPM---VSCPTCNIPICPTVSGYALYORHIKAMHESALLLEEQRQREQI 983  
Db 1806 QASETELAAGIIGSINDISGEPENFPAPPYPG-----ESQT 1842  
Qy 984 DLECRSSTSPCGTSKSPNRENEVLQAPHQLITNLPBGVRLPT---TRPTRPP---PPLI 1037  
Db 1037

Db 1843 DLQ-----PPAGA-----QALQPSSEGMETDEAVSGILETEAATBSSRPVPNAPD-- 1887  
Qy 1038 PSSKTTVASEKPSFTMGSGISQGTPTGTYLTSHN--QASYTOETPKPSVGSISLGLPROQE 1095  
Db 1888 PSAGPTDKE-----ARGNSSE-----TSHSVPEAKSGSKE-----VEVTLVRKDK 1927  
Qy 1096 SAKSATLPIYKO-----EEFSPRSONSOPEGLLVRQAQHEGVVRGTAGATQE---G 1142  
Db 1928 GRQKTRRRRRNTNKKVVAVPESHVESNOAQGES---PAANEGETTVOHPEAPQOEKQS 1984  
Qy 1143 SITRGTP-----TSKISVESIPSLRGSITQGTALPQOTGIPTEALYKGSISRMPIEDSSPE 1198  
Db 1985 EKPHSTPQOSCTDLS--KIPSTENS--SQEISVEERT--PTKASVPPDLP--PPQPAPV 2037  
Qy 1199 KGREEAASKGHVIEG-----KSGHILSYDNINKAREGTRSPRTAHEISLKRSTVESVEGN 1253  
Db 2038 DEEPPQARFRVHSIIIESDPVTPPSPDIPITPLSPVTAAKLSPVVA---SGGIPHQSPPTK 2094  
Qy 1254 IKQGMSMRESP--VSAPLEGLICRALPRGSPHSDLKE-----RTVLGSGSIMQGTPTATTE 1306  
Db 2095 VTEMITRQEEPPRAOSTP-----SPALPPDTKASDVDTSSSLRKLILMDPKYVSATSVTST 2149  
Qy 1307 SFEDGLKYPKOIK---RESPPIRAPEGAITKGKPYDGTITTIKEMGRSTHEIPRODILTOE 1363  
Db 2150 SVTTAIAEPVSAAPCLHEAPP-----PPVDSKKPLEEKTAAPPVTNNS--EIOASEVLVA 2202  
Qy 1364 SRKTPEVVOSTRPIIEGSGISOGTIKPD--NNSGOSAIKHNKVSILITGPKSLSRGMPPLEI 1422  
Db 2203 DKE--KVAPVTAPKITSVISR--MPVSDILENSQKITLAKPAPQTLTGLVSAULTGLNVSL 2259  
Qy 1423 VPEN--IKVVERGKYEDYKAGETVRSRHTSVVS--SGP--SVLRSTLHAPKAQLSPGIYDD 1478  
Db 2260 VPVNALKGPKVGSYTLK-----SLVSTPAGPVNVLKGPVNV-----LTGP 2300  
Qy 1479 TSARPTV-----SYONTMSRGSPMMNRT--SDVTIPPKNKSTNHERKSTLTP----- 1523  
Db 2301 VNVLTPVNTATVGTVNAAPGTVNAASAVNATASAVTAVTAGAVTAASGGVTATTCTVMA 2360  
Qy 1524 -----TORESIPAKSPV--PGVDPVVSHPSPDHPHGRSTA-----GVYKSHL 1564  
Db 2361 GAVTAPSTCKQORASANENSRFHGSPVIDDRPAD---AGSGAGLRVNTSEGVVLLSYS 2417  
Qy 1565 -----PTQLDPAMPFHRALDPAAAYL--FORQLS-----PTPGYPSOYOL 1603  
Db 2418 GOKTEGPORISAKI---SQIPPSAMDIEFOQSVSKQVRPDSVTASQPPSKG--POAPAG 2473  
Qy 1604 YAMENTQTLINDYITSOQMVNLRPDVARCLSPR---EQP-----LGLPYPATRG----- 1651  
Db 2474 YANVATHSTLV---LTAQ---TYNASPVISSYKADRPSEKPEPIHLSVSTPVTQGGTVK 2528  
Qy 1652 IIDLTNNMPT-----ILVPHPGGTSTPPMDRITYIPCTOITFPPRPVN--SASMSGHPT 1704  
Db 2529 LTQGINTPVVLVHNLVLTFSIVTNNKLDAPVLKTIETKVL---QANLGSITLPHHP 2595  
Qy 1705 HLAASAAEREREREREK---ERERERIAAASDLYLRPSEQPGRPCSGHGYVRSPPSP 1761  
Db 2586 ALPSKLPTENVHVPSPGSIADRTVSHLAAKLDH-----SPRSPGCPSPSFRASHPS- 2640  
Qy 1762 RTOETMLQORPSVF--OGTNGTSVITPLDPTAQLRIMPLPAGGPSISQGLPAS----- 1812  
Db 2641 STASTALSTNATVMAAGIPVQPFISLHP--EQSVIMP-----PHSITQTVSLSHLSQGEV 2695  
Qy 1813 RYNTAADALAAALVDAASAPOMDYSKTESKHEAARLEENLRSRSAASVSEQOOLEQKTL 1872  
Db 2696 RMNT-----PTLPSITYSIRPEALH-----SPRAPLQPOQ----- 2725  
Qy 1873 VEKRSVOCLYTSSAPFGKQPOHSSVYVSEAGKDGPPPKSRVEEELRTRGKTTITANF 1932  
Db 2726 IEVRAPQRASTPQAPAGVP-----ALASQHP-----EEVHYHLPVARATAPV 2770  
Qy 1933 IDVLTITQIASDKDARERGSSDSSSLSSHRYETPSDALEVISP-----ASSP----- 1982  
Db 2771 QSEVLVMQ-----SEYRLHPYTVPRDRVIMVHPHVTAVSEQPRAADG 2812













Db 724 RSLMSS--PEDLTKDFEELKAAEIDVARDIKPOLELIEDEEKLTEPGEAYVIQKETE 781  
QY 758 ETAPSTSPSLAVPSTK-----PADESVEVETQVNDSI----- 788  
Db 782 SKGSAESPDEGITTTEGECEQTEPELEPVKEGVGDIEKFEDEGAGFESSAGYEE 841  
QY 789 --SAETAQMDVDOOQERHSAEAGSYCDPPATKADSVDEVVRVPHENHASKVGDNTKERDL 846  
Db 842 KAEYEEAEPEDEEDNVSASAKHSPTDEDEIAKAEADVHIKRESVASGDDRAEEDM 901  
QY 847 DRASEKVEPRDEDLVAAQINQAQPERQSDNDSSATCSADVEDVGEPERQRMFMDSKPS 906  
Db 902 DEALEKGE-----AEQSEGEDEEEDKAEDAREE--DHEPDKTE--AEDYVMA 945  
QY 907 LLNPTGSLVSS--PLKPNPLDLPOLQHRAAVIPPMVSCTPCNIPICTPVSGVYALYQRH 963  
Db 946 VVDAEAGVTEQDYFLGTPTAKQGVQS-----PSREPAS----- 981  
QY 964 IKAMHESALLEEQROREQIDLECRSSTPCGTSKSPNREWEVLQAPAPHQILITNLPEGVR 1023  
Db 982 --SIHDETL-----PGGSESEATADEENRE--DQEEFTATSGYTOSTI 1022  
QY 1024 LPTRPT-----RPPPLIIPSSKTTIVASEKPS--FINGGSIISQGTPTGYLTSHNQASYTQ 1076  
Db 1023 EISEPTPMDEMSTPRVMTDETNEETESPQEFV-----NITKYESSLYSQ 1070  
QY 1077 ETPKPSVGSISLGLP-----RQGESAKSATLP--YIKOEFSRPSQNSQPEGLLV 1124  
Db 1071 EYKPVVASFN--GLSDGSKDTATDGRYNASASTISPPSMEEDKFKSALRD-----AY 1124  
QY 1125 RAQHEGVVGTAGAIQSGITRGTPTTSKISVESIP-----SLRGSITQGTPTAL 1172  
Db 1125 RPEETDKTGAELDKOVSDERLSPAKSPSLSPSPSPPIEKTPLGERSVNFSLTPNEIKA 1184  
QY 1173 POTGIPTALVKGISIRMPED--SSPEKGREEAASGHVIEKSGHILSYD---NIKN 1227  
Db 1185 SAEGEAT--AVVSPGVQAVVEHCASPEEKTLEVVSQSQV--TGSAGHTPYIQSPTEKS 1242  
QY 1228 AREGTRSPRAHEISLRKSV--ESVEGNIKQCMRMREPVSAPLGLICRALPRGSPHSDL 1286  
Db 1243 SHLPEVTENAAQVPVFETEAKDENERSISIPMDEPV-----POSES 1286  
QY 1287 KERVLSGSIQMGTPRATES--FEDGLUKYPKQI---KRESPPIRAFEGAITKGPYDGIT 1342  
Db 1287 PIEKVLSP--LRSPPLIGSESAYEDFLSADDKALGRSESP----FECKNGKQGFSDKES 1340  
QY 1343 TIKEMGRSHE-----IP--RQDILTOESKRTPEVV--QSTRPIIEGSI--SQGTPI 1388  
Db 1341 PVSOLTSOLYQDKOEKRAKRAFIPIKEDFSPEKKAASDAEIMSSQSALALDERKLGGDGSPT 1400  
QY 1389 KFDNNGSQAIAKHNKVSILITGPSKLSRGMPLLEIVPENIKVVERGKYEDVKAGETVRSRH 1448  
Db 1401 QVD--VSQFGFKEDTKMSISEGTVSDKSATP-----VDEGAEDTYSHMGCVASVS 1449  
QY 1449 TSVVSSGPSVLRSLTHBAPKAQSLPGIY-----DDT-----SARRTPVSYQNTMSR 1494  
Db 1450 TA-----SVATSSFEPTTDDVSPSLHAEVSGSPHSTVEDDSLSVSVQPTPTFOET--- 1500  
QY 1495 GSPMKNRTSDVTIPNKNSTNHERKSTLTPTQRESIPAKSPVPGVDVPPVSHSPFDPHGRS 1554  
Db 1501 -----EMSPSKE--ECPRPMSTSPDPSFKTAKRTPVOD-----HRSE 1537  
QY 1555 TAGEVYVWHLPTQLDPAHPFHR--ALDPAAAAAYLFORLSLPTPGYPS--QYOLYAMENTRQ 1611  
Db 1538 Q-----SSMSIEGQSPHSLAND-----FSRQ---SPDHTVCGAGMLHITENGPT 1581  
QY 1612 TILNDYITSOOMQVNLPRDVARGLSPREQLGLPYPATRGIDITJNPPITILVPHPGGTS 1671  
Db 1582 EV--DYSPSDIQDSSL---SHKIPPTPEP---SYTDNDLSELISVSQVEASPTSSAH 1632  
QY 1672 TTPMDRITYIPGTQITPPPPRYNSASMSPGPHLAAASAEERERERERERERERERERERER 1731  
Db 1633 TP-----SQIASPLOEDTLSDYVPPRDMSLYASLASAEKVSQSLGEKLPKSDISP 1682

QY 1732 ASSDLYLPGSEQPRGSGHYVRSPSPSVRTQETMLAQORPSVFQGTNGTGVITPLDPTA 1791  
Db 1683 -----LTPRESSP-----TSPGFSDSISGAKESTAAYQ-----TSSPPIDAAA 1722  
QY 1792 QLRIMPLPAGGPSISQGLPASRYNTAADALAALDAASAPQMD-----V 1836  
Db 1723 -----AEPYGFRS-----SMLFDTMQHHLALSRLDTTSSVEKDNKGKTPGDFNYAYQKP 1771  
QY 1837 SKTKES-----KHEAA-----RLEENLRSAASAVSEOOQOLEQKLEVEK 1875  
Db 1772 ESTTESPEDEEDYDYESHEKTIQAHDVGYYEYKERTIKSPCDSGSYSETTEKTKTPED 1831  
QY 1876 RSVQCLTSSAFFGKPOPHSHSVYVSEAGDKGPPKPSRYEELTRGKTKTITAAANFIDV 1935  
Db 1832 GGYCEITEKTRTPPEGGYS---YEISEKTRTPPEVSGYTYEKTERRRLD----- 1881  
QY 1936 IITROIASDKDAREGSSQSSSSLSLSHRYET-----PSDAIEVISPASSAPPQE 1987  
Db 1882 ---DISNGYDDTDGGHTLGD---CSYSYETTEKITSPPESESYSYETTTKTRSPDT 1933  
QY 1988 KLQIYQ--PEVVKANQANDPTROYEGPLHHRPOOESPS-----PQ 2027  
Db 1934 SAYCIETMEKITTKTPQAS---TYSIETSDRCYTPERKSPSEARQDVLCLVSSCFKHPK 1990  
QY 2028 QQLPPS---SOAEGM--GOVPRTHRLITLADHICQIITQDFAR-----NOV 2068  
Db 1991 TELSPSPINPNLWFAGEEP-----TESEERPLTQSGGAPPSPGGKQ 2034  
QY 2069 SSOTPOQPTSTFQNSALYSTVTRKTSNRYSPESQAQSVHHQRPGRSVSPENL--VD 2126  
Db 2035 GRQDETPTPTSVSEAPS-----QTDSDVPPETE-----ECP--SITADANLSD 2078  
QY 2127 KRSRGRCKSPERSHVS--SEPYEPISP--POVPVYHEKQDSLILLLSQGA----EPARQR 2179  
Db 2079 ESETIPTDKTVYKHMOPPPAPMQDRSPRHPDVSMDPEALAEQNLKALKKDLKEK 2138  
QY 2180 NDARSPGISYLSPLSFFTKLENTSPWVSKKOEIFRKLNSGGSDMAAQAOPGTEIFNLP 2239  
Db 2139 AKTKPG-----IKTKSSPVKK-----GDGKSKPSAASP-----KPG 2171  
QY 2240 AVTTSGSVSSRGHGFADPASNLGLEDIIR-----KALMGSDFKVEDHGVWMSQPMGV 2292  
Db 2172 ALKESSDKVS---VASPKKESVEKAMKTTTPEVKATRGEEKDKETKNAANASAKSV 2228  
QY 2293 V-----PCTANTS 2300  
Db 2229 KTATAGPCTTKTA 2241

RESULT 9

Q9NJ17

ID Q9NJ17 PRELIMINARY; PRT; 5476 AA.

AC Q9NJ17;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE SPLIT ENDS.

GN SPEN OR CG18497.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20157049; PubMed=10655223;

RA Rebay I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Lavery T.,

RA Suh C., Voas M., Williams A., Rubin G.M.

RT "A Genetic Screen for Novel Components of the Ras/Mitogen-Activated

RT Protein Kinase Signaling Pathway That Interact With the Yan Gene of

RT Drosophila Identifies split ends, a New RNA Recognition Motif-

RT Containing Protein.";







QY 1873 -----VEKRSVOCYLTSAFFSGKPPQSHSVVYS----- 1901  
DB 4657 MPQHPKKAIAAAETESYQAI--NSSIPSGGLPMHQTAAPATQKITGGVADAVSKALVDP 4714  
QY 1902 -----EACKDKG-----PPKRSYEEELTRGKTTITAAAN- 1931  
DB 4715 VTGVTGAMPQCKEKNLPAATAAPANSNEDGQAAPPQLQHQQOQHPOPPQOQANL 4774  
QY 1932 -FIDVITRQIASOKDAREGSSQSDSSSLSSSHRYETPDAIEVISPASSAPPOEKILQ 1990  
DB 4775 QINTLIPSLPNPTALGKSVOLETSAALL-----NKPVSVLVKGNA-----SQVIQ 4823  
QY 1991 TYQPEVVKANAENDPTRO-----YRPO-----LHH-----ESPPOQ 2029  
DB 4824 OQOQOIVA-----PAKQPIIQQONPLPTVLHHAQHTVRRPQPKAHVLNREKNIQOQ 4876  
QY 2030 LPPSSOAGMGQVPRTHLTIADHICOIITQDFARQ-----VSSQTPOQ----- 2075  
DB 4877 LPTTKQAVQ-----PPOH-----APHSGHMLITDTAGNOOLVOPQIIARHLOQQOHLQVNV 4928  
QY 2076 -PPTSTFQNSALYSTVTRTKTSNRYSPESAQSVHQRPGSRVS-----PENLVDKSR 2129  
DB 4929 PPTAHSPHSRI-----PSOQQO-----LPGCASISPOOQOPOTVVIKQA 4969  
QY 2130 GSRPGKSPERSHSEPEYIPSPQVPVYHEKQDSLILLSQSG-----AEPAEQRNDARS 2184  
DB 4970 AS--AAQPIQLHVSSKASVVPQO-----QQQLPPTSTSGPHLQAKP----- 5011  
QY 2185 PGSTISLPSFTFKLENTSPMKSKOE-IFRKLNSGGGSDM-----AAAQPGTEI-F 2236  
DB 5012 --NYSIAP---TVLTPTLPVQOQOQOHLKYKONNQKQAKQIOMPHGIMPTHGMLLQ 5066  
QY 2237 NLPAVTTSGSVSSRGHSPADPASNLGLDIIIRKALMGSDDKVEDHGVVMSQPMGVVPGT 2296  
DB 5067 KLPA-----HLOQOQOHL-NPSPPGKPNVL-----HGLQSGQ-----IMPGS 5104  
QY 2297 ANTSVVTSGTRREEDGSPHSGGVCK-----PKLISKSNSRKS 2337  
DB 5105 V-----GSPPPVSAALVLTAAQQVNSVVPVAGIRTAIPNISQSQPRVSPL 5150  
QY 2338 PIPGOGYLGE-----RPSVSVSHSGDY-----HROTPGWAMEDRPSSTGSGTFYPNP 2387  
DB 5151 VLP-PGISGVPPFDASLNDSTYRGVTSASDFMLYQHHLMRGGDYDDKMGSS-----PP 5202  
QY 2388 LTRMLSLSTPPTIACAPSAVNQAAPHQONRIWEREPAPLLSAQVETLSDS 2438  
DB 5203 LELRRPGSGPRTI-----AVPHS-----LQSPQRTAADS 5233

RESULT 11  
Q9Q2W2  
AC Q9Q2W2 PRELIMINARY; PRT: 3576 AA.  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE MSX2 INTERACTING NUCLEAR TARGET PROTEIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RN RC TISSUE=TESTIS;  
RA MEDLINE=9379811; PubMed=10451362;  
RA Newberry E.P., Latifi T., Towler D.A.;  
RT "The RRM domain of MINT, a novel msx2 binding protein, recognizes and  
RT regulates the rat osteocalcin promoter."  
RL Biochemistry 38:10678-10690(1999).  
DR EMBL; AF156529; AAD55931.1; -  
DR InterPro; IPR000504; -  
DR InterPro; IPR001993; -

DR Pfam; PF00076; Irm; 3.  
DR PROSITE; PS00215; MITOCH\_CARRIER; UNKNOWN\_1.  
DR SMART; SM00360; RRM; 1.  
SQ SEQUENCE 3576 AA; 390966 MW; 9B56855D8A8F38A8 CRC64;  
  
Query Match 3.8%; Score 484.5; DB 11; Length 3576;  
Best Local Similarity 19.9%; Pred. NO. 2e-17;  
Matches 492; Conservative 298; Mismatches 875; Indels 811; Gaps 118;  
  
QY 4 SCYGP-----NOGAFSTEQSRYPHPSVQYTFNTRHQEFAYPDYRSSHLEVSQASQ 55  
DB 1500 SDFPGRLYGROASEGANSTSDSVOEP-----VVLFRHSRMELTR----- 1539  
QY 56 LIQQOQOQOQLRRRPLLSEFHFGSDRQERRTSYEP-FHP-GPSVPVDHDSLESKPRLEQ 113  
DB 1540 -MQQKEKEK-----DQPKAEKQEEPEHPTKTEP-----AATKEPEPKA 1580  
QY 114 VDSHFQRYSAVLPLVHPLPEGLRASADAKDPAGFGKHEAPSSPIGQ----- 163  
DB 1581 PVSAGLPATITVV-TPEPASSAPEKAEAEAAEAPSPAGEKPAEPAPVSEETKLVSPEVSV 1639  
QY 164 -----PGDDONAS-----PSKLSKEELIQSDRDVRE-----IAKVE--- 196  
DB 1640 PVEQPRQSDVPPGEDSRDSQDSALAPSAQESAATDAVPCVNAEPLTPGTTVSVQESSV 1699  
QY 197 -----QOILKKKKQOOLEE-----EAAKPPPEPEKPVSP 226  
DB 1700 DPKSSPQPLSKLTQSEAEKGVEKPDTPSTPDTATONAGVASEAQPPASEDEANP 1759  
QY 227 PVEQKHSRIVQIYDENRKKAEAAHKIFEGLGPKVPELPLYNQPSDTKYVYHENIKTNQVMR 286  
DB 1760 PVAADR-----KTNKSRKSTSVQAAAAVVEKPVTR-----KSERIDR 1799  
QY 287 KKLILFFRRNHARKQREKICQRYDQLMEAWKKVDRIENN-----PRRKA 333  
DB 1800 EKLKRSSSPRGEAQKLELK-----MEA--EKITTASKSSGGDTEHPSPSLPSRS 1849  
QY 334 KESKTREY-----EKOPPEIRKQREOQERFORVGORGAGLSATIAKSEHISIIDGL 387  
DB 1850 RRRNVRSYATWTDHESRP--AKEPVEQPVTR-----KLERELQE----- 1890  
QY 388 SEQENNEKOMRQLSYIPPMFMDAEORRYKFINMGLMEDPMKYKDRQFMNVTWTDHEKEI 447  
DB 1891 -----AVPP-----TTPRRGR-----PP-----KTRRAEDGGEHERK- 1919  
QY 448 FKDFIQHPKNEGLIASYLERKSVDCVLYYYLTFKKNYKALVRRNYGKRGRRNQOIA- 506  
DB 1920 -----EPAETP-----RPAEGWRSPRSQKSAA 1941  
QY 507 -----RPSOEKVEEKEEDKAEKTEKKEBEK-----KDEEKEDEKSKENTKDKIDG 556  
DB 1942 AAGPQKGRNEQKVEAAAEAGAAQASTREGNPKSGEREAAASEPKRRDRDSTDSGPD 2001  
QY 557 TAEETEEREQATPRGRTANSQGRKGRITRSMTNEAA-----AASAAAAATEEPP 609  
DB 2002 FPVEVLERKPP-----EKTYKSK-RGRARSTRSGMDRAAHQHSLEMAARAAGQAADKEAGP 2056  
QY 610 PLPPPE-----PISTEPVETSRWTEEMEYAKKGLVHGHRNWAIAIKWGTG 657  
DB 2057 AAASPOESESQKSGSSPQALNANPADPDRAEESAS-----ASTAPPEGTQ 2104  
QY 658 SEACKNFFYFKRRHNDNLQOQHKQTSRKPREERDVDSQESVASTVSAQEDEDIAS 717  
DB 2105 LARQIE-----LEQAVQNIKLP--EPSAAAASKGTATATA-----TAA 2141  
QY 718 NEENPEPSEYEAVKPSDSDSPENATSGNTEPAVELETTETAPSTSPSLAVSPKPAED 777  
DB 2142 SEEPAPHGHPAKQASE--TELAAGISIIISDASGEFENFAPSVP-----PGSQTHPR 2195  
QY 778 ESVETQVNDSTISA--ETAEQMDVDQOQHSABEGSVCDPPPA-----TKADSVDEVVRVP 829  
DB 2196 EGMEPGLHEASGILETC-----TATESAPQVSALDPPGEGSADTKETRGNSGD---SVQ 2247

QY 830 ENHASKVEDNTKRDIDRASEKVEPR-----DEDLVAAQIINAQRPQSDNDSDATCS 884  
DB 2248 EAKGSKAEVTPRK---DKGROKTRRRKRANKKVAITETRASEAE-QTQSESPA--- 2300  
QY 885 ADEDVDGEPERQRMFMDKPSLLNPT--GSILVSSPLKPNLDLPQLQHRRAAVIPPMVSC 943  
DB 2301 AEEATAATPEA---POEEKSEKPPSPAECTFDPKSTPPAESLSQENSA---EK 2350  
QY 944 TPCNPICPTPSGVYALYORHKAMHESALLEEQRORQOIID--LECRSSTSPCGTSKSPN 1001  
DB 2351 TPCKAPV-LPA-----LPLSQPALMDGQPAQKVIHSIESDPVTPPSDSGIPP- 2399  
QY 1002 REWEVLQAPAHQIINLNGVRLPTRTRPPPLI-----PSSKT-----T 1043  
DB 2400 -----PTIPLVTIAKLPPV-IFGVPHQSPPPKVTWITROEPPRAQSTPSPALPDDT 2452  
QY 1044 VASE-----KPSFMGSGISOGTGTGYLTLSHNOASYTOETPKPSVSGSISLGL 1090  
DB 2453 KASMDTSSSLRKLMDPKYVSATGVTSTVTTAIAEPVSAQCLOEAPAPCD----- 2506  
QY 1091 PROQ--ESAKSATPYIKOEESPRSONSOPEGLLVRAQHEGVGRGTAGATQEGSITRGT 1148  
DB 2507 PKHPPLEGVSAAVP-----NADTQASEVPVAADKEKVPVIAPKIT--SVTSRM 2554  
QY 1149 PTSKISVESIPSLRGSITOGTPALPO--TGIPTEALVKGSI SRMPIED--SSPEKGREAA 1205  
DB 2555 PVS-IDLEN---SOKITLAKPA-POITLGLVSAITGLVNLVLPVNAKLPVKG-SVAT 2607  
QY 1206 SKGHVYIEGKSGHILSYDNINAREGTRSPRTAHEISLKRSYESVEGNIKOGMSRESVP 1265  
DB 2608 LKG--LVSTPAGPVLNLLKGPVNLVTPGVNLVTPVSATGVTNAAPGPVTAACGVTAITG 2665  
QY 1266 SAPLEGLICRALPGSPHSLDKERIVL---SGSIMOGTPRATTESFEDGLKYPKOIKRES 1322  
DB 2666 TAAVTGAVTAPAAKGQRASNEHSRFPGMSVIDDRPADTGG--AGLRVNTS---EG 2720  
QY 1323 PPIRAFEAGITKGPYDGIITTIKEMGRSIEHP---RODILTQSRKTPPEV-VOSTREPII 1378  
DB 2721 VVLLSYSGOKTEGP-----QRISAKISQIPPASAMDI EFQSVSKSQVKA DSITPTQ 2772  
QY 1379 EGSISOGTPIKFDNNGSOSAIKHNKVSILITGP--SKLSRGMPPLIEIVDENKIVVERGYE 1436  
DB 2773 SAPKGPQPSAFANVAHASTLVLTAQTYNASPVISSTVKTDRPSLE-KPEPIHL---SVST 2828  
QY 1437 DVKAGETVRSRHTSVVSSG---PSVLRSTLHEAPKAQLSPGIDYDTSARRTPVSVQNTMS 1493  
DB 2829 PVTGGTVK-----VLTQGIWTPPVLL---VHN--QLVLTPTSIVTNKKLADPVT----- 2872  
QY 1494 RGSPMNRSTSDVTIPPNKSTNHERKSTLTPQRESIPAKSPVPVGVDPVYVSHSPDFPHRG 1553  
DB 2873 -----LKIEITKVLOPANLG-----PTLTHPPALPSKLP AE-----VNHVPSGP--- 2912  
QY 1554 STAGEVYVSHLPTQLDPAMPFHRALDPAAAYLFQORLSPTPG-----YPSQOLYAME 1607  
DB 2913 STPADRTIAHLAT---PKPDTHS-----PRPTGTPPGLFPRPCHPS----- 2951  
QY 1608 NTRQITLNDYITSOOMQVNLDPDVARGLSPREQLGLPYPATRGIIDLNTNMPILYVPH 1667  
DB 2952 -----TISTALSTNATVNLAAAGI-PVQFISSIHPEQSVIMPHSITQVSLGHL 3000  
QY 1668 G-----GTSTPPMDRITYPTGQITFPFRPNYSASNGHPHPLAAASAEEREREREKE 1723  
DB 3001 SQGEVRNSTPLPSTISYIRETILSPR-----APLOQOUIEARAQ----- 3042  
QY 1724 RERERIAAASDLVRCGSEOPGRGSHGVYVRSVSPSVRTQETMLQORPSVFOGTNGTSV 1783  
DB 3043 -----RVGTPOPATTTG-----VPALATQHPPEEE----- 3066  
QY 1784 ITPLDPTAQLRIMPLPAGGPPSISOGLPASRYNTAADALAAALVDAASAPQMDVSKTESK 1843  
DB 3067 -----VHYHLPVAR-----AAAP----- 3079

QY 1844 HEAALEENLRSAASEQOOLEKOTLEVEKRSVOCLYTSSAPSGKPOPHSSVYVSEA 1903  
DB 3080 -----VQSEVLVMQSEYRLHPYTV---PROVRIMV-----HPHVTAV----- 3113  
QY 1904 GKDKGPPPKSRYEEBELTRGKTTITAAFNFDVITITQIASDKDARERGSSQSSSSLS 1963  
DB 3114 -----SEQPRATEGVVVPANKAPQLVKEAVKTSDAK----- 3147  
QY 1964 HRYETPSDA-LEVISPASSAPP---QEKLOTYQEVVVKANQAEENDPTROIEGPHHYRP 2019  
DB 3148 ---AVPAPVPVFPVPTPAPPHPGEARILT----- 3176  
QY 2020 QOESPSQOQLPPSSQAEAGMGOVPRTHRLITLADHICOI-----TODFARNVSS-OT 2072  
DB 3177 -----VTPSQLOGLPLTP-----VVTHGVOIVHSSGELFOEYRYGVRTTHA 3221  
QY 2073 POQPPTSTFQNSPALVSTPVRTKTSNKNRSPESQ-AQSVHHQRPQ-----SRVSPENLV 2125  
DB 3222 PAOQLTHT-QPPVASSISLASRTKTSQAQVPEGEPLQTSQAQAPAPSTQATQPIPPAPPC 3280  
QY 2126 DKSRGSRPKSPERS--HVSE-----PYEPISPPQVVPVHEKQDSLLLS 2169  
DB 3281 QPSOLSQAQPSKIPQVSOEAKGTGTGVEQTRLPATPNRSEP--H-----A 3329  
QY 2170 QRGAEPAEORNDAR-SPGSI SYLPSFTFKLENTSPMVKSKKQEIFRKLNSGSGGSDMAA 2228  
DB 3330 OLQAPVETAQAPNPSVSVSMKFDLPSPSSQA-----APKQPLFVPANS--GPSTPPGL 3383  
QY 2229 AOPGTEIFNLPAVTS 2244  
DB 3384 ALPHAEQVAPAPKQESS 3399  
RESULT 12  
Q9NHNI  
ID Q9NHNI PRELIMINARY; PRT; 5554 AA.  
AC Q9NHNI;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE SPLIT ENDS LONG ISOFORM.  
GN SPEN OR CG18497.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kuang B., Wu S., Shin Y.-A., Luo L., Kolodziej P.;  
RT "Split ends encodes large nuclear proteins that regulate neuronal cell  
fate and axon extension in the Drosophila embryo.";  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF221715; AAF34661.1;  
DR FlyBase; FBgn0016977; spen.  
DR InterPro; IPR000504;  
SQ SEQUENCE 5554 AA; 599188 MW; 4037E27833D0C622 CRC64;

Query Match 3.8%; Score 484.5; DB 5; Length 5554;  
Best Local Similarity 18.1%; Pred. No. 3.7e-17;  
Matches 556; Conservative 389; Mismatches 1050; Indels 1085; Gaps 141;

QY 36 QEFVAVPD--YRSSHLEVSQASQLLOQQ---QQQLRRRPSLLSEFPGSDRPQERR----- 86  
DB 2583 QSQAPDLTIKQEH-PIAPAQEI KREQLSDEQKFKR-----HDSNSSIEERKLKTE 2634  
QY 87 -----TSVEPHFGP-----SPVDHDSLESKRPRLEOVSDSHFQVRYSAAVLPLV- 130  
DB 2635 REIKTELCDFYNSSEYITTKLKEYSPETRKHKRKLKSSSTAD---TSAATPLVM 2691  
QY 131 HPLPEGL---RASADAKK--DPAFGKGHEAPSPISGQCGDDQNAQSPSKLSKEELIQSM 185



Db 2692 TPLTPSIFDVHSSSECKTKFNFDDDKTECCSIPLE--ISAGERRKHKKERKKKKLNNM 2750  
Qy 186 -----DRVDREIAKVEQOILKKKKOOLEEE-----AAKPPPE 220  
Db 2751 TEATVPNSPTTNDTSSEKSEERHRLAKSKKSKMDNSCNTKIYNSSGAHPSTSPSLPA 2810  
Qy 221 KPVSPPV-----EQHRISIVQIYDENKKKABEAHIFEGL-----GPKVE--L 263  
Db 2811 TPTSAPTAOTS KRGEKMEFIFGIISDEESQFPPEAETNKNKIIPSSVSTGTIVSAAAL 2870  
Qy 264 PLYNQPSDTK-----VYHENIKTNQVMRKLLILFFKRRNHARKQREOKICQRYD 312  
Db 2871 QTYQEPSTPNKNEAAHIQTVHEPEOQOQLERSLSSGSSSSSHADRERH----- 2922  
Qy 313 QLMFAWEKVDRIENNRPRKAKESKREYIEKQFPEIRKQREQORFORVQORGAGLSAT 372  
Db 2923 -----RREKREKKRE-----KSQREQQNOI----- 2943  
Qy 373 IARSEHEISEILDSEQENNEKQMRQLSVIPPMFMDAEQRRVKFIPNNGLMEDPMKYK 432  
Db 2944 -----HOKSSKVKVDDNS-----VDMDEAGRALE-----AQLMSD----- 2976  
Qy 433 DRQFMNWTDEHEKTEFKDKETQHPKNFGLIASYLERKSVPCVLYYLTKNENYKALVR 492  
Db 2977 -----FDTKPISEATPTAATY--RSDMTDVFR-----SDNE----- 3009  
Qy 493 RNYGKRRGRNOQIARPSOEKVEKEEDKAEKTEKKEDEKDEKDEKSKENTKEKD 552  
Db 3010 -----NNSVDMTKGVKSEQEQHKS--DKKKKKKSKKEKQEKLLQOQRRSLP 3058  
Qy 553 KIDGTAETEERQATPRGRTANSQGRKKGRITR----- 587  
Db 3059 NVASTSS-----APPTP--GLTVNVQAASKHADLQDAKHISPPVCKPSPSLPCLIGDD 3112  
Qy 588 -----SMTNEAASAAAAATEEP--PPLPPP-----EPISTEPVETSRWTE 630  
Db 3113 DDDALHTPKAKPTTPSSGNDGLTPSRKPKLIPIPKTPIANSSTLSTQSAETPVSSG 3172  
Qy 631 BEMEYAKKGLVHGRNMAAIKRVG-----TKSBAQCK----- 663  
Db 3173 TVISSALATTPSTSTAAGVSAAPGLDNPSTPSASAQCKKESFIPGPDQLDDRTSESAB 3232  
Qy 664 -NFYFNKRRINLNLQOQKQTSRPREDRVSCQESVAS--TVSAQDE----- 712  
Db 3233 QSISAENSTSLDNIADPRKIPVAPSPRATKPLDKLEESKSRVTISOEETESAVSALLG 3292  
Qy 713 -----DIEASNEE-----NPDESEVAVKPSDESPENATSRGNT 748  
Db 3293 ESFGTSTTDSLDCGDEMSSVNELEPTLVIAEPDEEAALAAKAIETAGEPASTL--EE 3350  
Qy 749 PAVELEPTTETAPSTPSLAVPSTKPAE-----DESVEQVND--SISAETA----- 794  
Db 3351 P--EMEPERAEPPDPEAETESPPVVEVLDPEELNKAQVSLKHEDMMDIKADTPQSERD 3408  
Qy 795 -QMDVDOQESHAEGSCVDDPPATKAQSDVVEVRVNPENHASKVEGDNTKRDLDRASEKV 853  
Db 3409 LQIDTDTEENPDE-----ADSSGSLKIDET-----VQSSSPKESISNNSP-- 3450  
Qy 854 EPRDLDLVAQOIAQRPEDOSNDSSATCSADEVDGEPERQRMFMDSKPSLLNPTGS 913  
Db 3451 TPRETANIDPNVESQ--PKLSNESTP-----QPSVITKFLPDLTPKTV--PAG- 3495  
Qy 914 ILVSSPLKPNPLDLOLQRAAVIIPPMVSCVTPCNIPIGTPVSGYALYQRIKAMHESALL 973  
Db 3496 -LPPSPVKIEPTTISKLQ--PLVQPVQTVLPAPHSSTGSGISANSVINLDSNVISSC-- 3550  
Qy 974 BEQRQREQIDLECRSTSCGTSKSPNREWEVLQAPHQILTNLPEGVRLPTRTRTPR- 1032  
Db 3551 -----SNTSASATASASASISFGSPTASQ--NAMPQASPPKQGPITPQ 3592  
Qy 1033 -----PPPLIPSSKTTVASEKPSFTMG--GSIQGTGCTVLTSHNQASYTOETPKP 1081  
Db 3593 QAIQTQSLIQPPTISIPQTPHFAVQPMVLSPQSHHPQPGTYMVGIRAPSPHLSHP 3652

Qy 1082 S-----VGSIS--LGLPROOESAKSATLPYIKOEE-----FSFRSONSOPEGLLVRAQHE 1129  
Db 3653 GRGVAQSRVLGQLSPVGRPMVSPQPOQVOQTOQHALLITSPQSSNISP----- 3702  
Qy 1130 GVVRTAGAIQBGSTTRGTPTSKIS-----VSIPSRLG--SITQGTTPALPORTGIPT 1180  
Db 3703 -LASPTTRVL-----SSNSPTTSKVNQYQPRNQVQPOQSPKSAVEQVTPQL--MTIPLQ 3756  
Qy 1181 ALVKGSI SRMPLEDSSPEKREEAASKGHVYIEGKSGHLSLYDNK--NAREGTRSPRTAH 1239  
Db 3757 KMTPIQVPHPTIILSKVVTVQPOQATQSOVASSPPLGSLPPHKNVHLNAHQNOQPOVTA 3816  
Qy 1240 EISLKRSEYVESGNTKQGMMSREPSVAPLEGLICRALPRGSPHSDKLRKRTVLSSIMOG 1299  
Db 3817 KMTAHOHQHMQQFMHQMIQRO-----QHMQOQOLHGOSOOITSAPOQHMQ 3864  
Qy 1300 TPRATTESFEDGLKYPKO-----IKRESPIRAFEAGITKGPYDGIITIKEMGRSI 1351  
Db 3865 OHQAQOQOHHNQHLNQOLHAQQHTOKHOAQOOFNQOIQ-----QHOSQOQ 3913  
Qy 1352 HEIPRODILTQE-----SRKTPVVQSTRPIIEGSIQGTPIKFDN 1392  
Db 3914 HOVOQNOAQOQOHLSSQOQHQQOQLOQLOQIKLQOHHGPOQOQKSPQGVGH 3973  
Qy 1393 NSGOSAI-----KHNKVSLLTG-----PSKLSRGMPPLEIVPENIKVVERGYEDVKAGE 1442  
Db 3974 LGGSTISFASOHNQOLPARGVPOOQHPOOLSHSP-----CKNTLVSVNOGVQPPAILTR 4030  
Qy 1443 TVRSRHTSVVSGPSVLSTLHEAPKAOLSPGIYDDTISARTTPVS--YQNT----- 1491  
Db 4031 VGSHSQPNQOQOLPHQOSSGHPHQQLSSPGA--NLQTPLVNIQNTPKIIVQOHIV 4087  
Qy 1492 -MSRGSPMMNRTSDVTIPPNKSTNHERKSTLTPTTORESIPAKS-----PV 1535  
Db 4088 AQNVPPPTQOTGNAIHPQNOG-----KDSTPPGHVTEPPAMSAOKTSESVIRTPPT 4142  
Qy 1536 PGVDPVVSHPSPDPHRRGSTAGEV-----YWSHLPTQ 1567  
Db 4143 TGL-AVIS-----ANTVGSLLTEENLIKISQKODELIEQDSKEVSDSYNAKEVN 4192  
Qy 1568 LDPAMPFRALDPAAYVLFORQL-----SPTPGYPSOYQLYAMENTRTI--L 1614  
Db 4193 IDSVI--KLDTPPLASKDAKRAVEMQAIAPAPINPQPNQSGMAQETALPTTSMVNS 4249  
Qy 1615 NDYITSOQMOVNLRDVARGLSPREQPLGLPYPARG-----IIDLTNMPPT 1661  
Db 4250 NDHDTEDTETRQUPPAKPPITPTVGRPPGGSAGRKGRPGAKKVGFPPLNSVTAAPPG 4309  
Qy 1662 I--LVPHPG--GTST-----PP-----MDRIT 1679  
Db 4310 VDSLVOQDNGVQTRLRKPVTA PVTGRKGPPRPNLLQOQLOQOQLOQDIOQKMEWT 4369  
Qy 1680 YIPGTQITFPFPNPNASMSGPHTHLAAAASAEEREREKERERERIAA---ASSDL 1736  
Db 4370 --SATSTPLPTPI-----PTSSVLTAAEKARNQALTAQOEQNAQVSGVTGDI 4418  
Qy 1737 Y--LRPGSEQCPGSGHYVRSPSPSVRTQE-----TMOQRPVS-----FGT 1778  
Db 4419 YEFHEDGGEEP--KPKT---ISSVAPSAEDQRPRLILITNKTPQSIKNISEMBQTOOQOQ 4474  
Qy 1779 NGTSVITPLDPTA-----QLRIMPLPAG--GPSISQ 1807  
Db 4475 QOSEVISNTDPIGGDNSESCNTRKSRRLQEKEDRSTVDDIIEDVVVNTNTPTGTGPHLPK 4534  
Qy 1808 GLPASRYNTAADALAALVDA-----AASAPQMDVSKTKESK---HEAARLEENLRSAA 1859  
Db 4535 GAQTPPRSGRNAQAKTDAVQIINAVGRPR-----RSKDRKTIGEQTANLIEEVTASNAT 4590  
Qy 1860 V-----SEOOOLEQKLE----- 1872  
Db 4591 VAASHLAPPEGAGVESHVPQDLDAKEVEPVSVVTPISTPAPVSVAAPVTVPVAMVPVKPT 4650

```
QY 1873 -----VEKRSVOCLTSSAFPSGKPKQPHSSVVYS----- 1901
Db 4651 MPOHPKKKAIAAEIESYQAI--NSSIPSGGLPMTAAPTAKITGKVADAVSKALVDP 4708
QY 1902 -----EAGKDKG----- 1931
Db 4709 VTGVTITAGMPQCKEGNLPAAATAAAPSNSNEDGAAPPQLOHQOQOQHPQPPQOQANL 4768
QY 1932 -FIDVITITQIASDKDAREGSSQSSSSLSHRYPSPDAIEVISPASSPAPQEKIQ 1990
Db 4769 QINTWILISGLPNPITALGKSVQLETSAAALL-----NRPVSVLVKGNA-----SQVIQ 4817
QY 1991 TVQPEVVKANOENDPTRO-----YEGP-----LHH-----YRPOQ-----ESPSPQOQ 2029
Db 4818 QOQOQIVA-----PAKQPIILOQNPPLTVLHHAQHTTVRPPQKAKAVLNREKNIQOQ 4870
QY 2030 LPPSSQAEQMGVPTTHRLITLADHICQIITODFARNQ-----VSSQTPQO----- 2075
Db 4871 LPTTKQAVAQ---PPQH-----APHSGHMLLTDTAGNQQLVQPIIARHLQOQOHLQVNV 4922
QY 2076 -PPTTFONSPALVSTPVRTKTSNRYSPESQAOQSVHHQRPGRSVS-----PENLVDKSR 2129
Db 4923 PPTTAHSPSPRI-----PSQOQO-----LGPQASISPOQOQPTVVIKQA 4963
QY 2130 GSRPKSPERSHVSSEYPIESPQVYVHEKQDSLILLSQSG-----AEPAEQRNDARS 2184
Db 4964 AS--AAQPOILHWSSKASVVPQO-----QQQLPPTSTGPHLQAKP----- 5005
QY 2185 PCSIYLSFFTKLENTSPWYKSKQOE-IFRKLNSGGGSDM-----AAAPQTEI-F 2236
Db 5006 --NYSYAP---TVLPPTLPFAVQOQOQOQHYKONNQKQAQIQMPHPHIMTPHGMLLQO 5060
QY 2237 NLPAVTTSGSVSSRGHSTADPASNLGLEDIIRKALMGSDFKVEDHGVVMSQPMGVVPGT 2296
Db 5061 KLPA-----HLQPOQHOL-NPSPPGKPNVL-----HGLQSGO---IMPGS 5098
QY 2297 ANTSVVTSGTTRREGDSPSHGGVCK-----PKLISKNSNRKSKS 2337
Db 5099 V-----GSPPPVSAVLKTAQOQVNSVPVAGIRTAIPNISQSPRVSPL 5144
QY 2338 PIPGOGYLGTBRPSSVSVSEGDY-----HRQ-----TP----- 2367
Db 5145 VLP-PGISGV---PPPDASLHDLGAVSGRRQTQSPPPAHQQAQSPITPNDSTYRGVTASRDF 5201
QY 2368 -----GNAWEDRPSSTGTSQPPYPLNPLMRLLSPTPTIACAPSANVCAAPHOQNR 2418
Db 5202 MLYQHHLMRGGDYDKMGSS-----PPELRLRPGSGPPRTI-----AVPHS----- 5242
QY 2419 IWREPAPLLSAQYETLSDS 2438
Db 5243 -----LQSPQDRTAADS 5254

RESULT 13
Q9VPL1
ID Q9VPL1 PRELIMINARY; PRT; 5560 AA.
AC Q9VPL1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SPEN PROTEIN.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
```

```
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abrill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bock J., Brokstein P., Brottier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,
Zheng X.H., Zhong F.N., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of Drosophila melanogaster."
Science 287:2185-2195(2000).
DR EMBL; AE003590; AAF51535.2;
DR Flybase; FBgn0016977; spen.
DR InterPro; IPR000504;
DR Pfam; PF00076; irm; 3.
DR
SQ SEQUENCE 5560 AA; 599996 MW; 58F19621AF40D2A8 CRC64;
```

## Query Match

3.8%; Score 484.5; DB 5; Length 5560;

Best Local Similarity 18.18; Pred. No. 3.7e-17;

Matches 556; Conservative 389; Mismatches 1050; Indels 1085; Gaps 141;

```
QY 36 QEFAPVD--YRSSHLEVSQASQLLQOQ---QQQLRRRPSLLSEFHPGSDRQERR--- 86
Db 2589 QSQPAPDLTIKQEH-PIAPAQEIKRQSLDEEQFKSR-----HDSNSIEERKLKTE 2640
QY 87 -----TSYEPFHPGP-----SPVDHDSLSKRPRLEQVSDSHFORVSAAVPLV- 130
Db 2641 REIKTELGDFNYSSEYTYTGKLEYSPETRKKHKKRKLKSSSTAD---TSAQTPLV 2697
QY 131 HPLPEGL---RASADAKK--DPAFGGKHEAPSPISGPGCGDDQNASPKLSKRELQISM 185
Db 2698 TPLPSPIDVHSSECKTKYFDNDDLKTECSSIPLE-ISAGERKKHKKERKKEKKLKNM 2756
QY 186 -----DRVDREIAKVEQOILKKKKQOOLEE-----AAKPPPE 220
Db 2757 TEATVPNSPTTNDTSSEKLSKEERHLLKSKSKSDMNSCNTKIYNSGSAHPSTSPSLPA 2816
QY 221 KPVSPPPV-----EQKHSRIVOLIYDENRKKAEAAHKIFEGL-----GPKVE--L 263
Db 2817 TPTSAPSTAQTSKRGEKMEFIFGIISDEESQPEQAETNKDIIPSVSTGPTIVSAAL 2876
QY 264 PLYNQPSDTK-----VYHENIKTNQVMRKKLILFFKRNHARKQEQKICQRYD 312
Db 2877 QTYKQEPSTPNKNEEAHIQITVHEPQQOQLERSLSGGSSSSSHADREH----- 2928
QY 313 QLMDEAWKVKVDRIENPRRKAKESKTREYEQKPFETRKQEQORFORVQORGAGLSAT 372
Db 313 QLMDEAWKVKVDRIENPRRKAKESKTREYEQKPFETRKQEQORFORVQORGAGLSAT 372
```







RA Sulston J.E., Waterston R.;  
RT "Toward a complete human genome sequence.";  
RL Genome Res. 8:1097-1108(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kraemer J., Wollam C., Wohldmann P., McGrane B.;  
RT "The sequence of Homo sapiens PAC clone DJ0784G16.";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Waterston R.H.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Waterston R.H.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Waterston R.H.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC004886; AAD21789.1; -;  
DR InterPro; IPR002965; -;  
DR PRINTS; PR01217; PRICEXTENS.  
FT NON\_TER 1  
ET NON\_TER 3394 3394  
SQ SEQUENCE 3394 AA; 375558 MW; 1979C29AE9B06B73 CRC64;

Query Match 3.8%; Score 480; DB 4; Length 3394;

Best Local Similarity 17.9%; Pred. No. 3.3e-17;  
Matches 553; Conservative 396; Mismatches 1084; Indels 1056; Gaps 128;

QY 35 QOEFVAPYRSHLE---VSQASQLQOQOQOQLRRPSSLSEPHGSDRPPQERTSYE 90  
DB 88 OBESEKLEKASALQKPLPEEKLIPEEKIRSEKPLLEKKP---TPEDKLLPE 144  
QY 91 PFHPGSPVDHDSLESKPRLEOVSDSHFORVSAVLP-----LVHPLPEGIRA 139  
DB 145 AKTSAPEQKHDLKQ---VQIAE---EKLEGRVAPKTVQEGKQPTKMEGLSPGTPQ 197  
QY 140 SADAKDPAPFGKHEAPSSPIGQP-----CGDQNASPSKL-----SKEELIQ 183  
DB 198 SLPKEDDTTKTIKEQPQPCTAKPDQEKEDKSDTSSQPKSPQGLSDTGYSDGISS 257  
QY 184 SMDRVREIAKVEQOILKLLKQOOLEEAAKPPPEKPPVPPVQKHSIVQI----- 238  
DB 258 SLGEIPLSIPTDEKIDILKGLKKSFSQESS-----PSSPSDLAKLESTVLSLEAQA 309  
QY 239 --IYDENKKAEEAHKIFEGLPKVELPLYNOPSDTK-----VYHENIKTNQV 284  
DB 310 STLADEKSEKTPQHEV---SPE-----OPKDEKTSQSLSETLEITSEEEIKESQE 358  
QY 285 MRKLLILFFKR-----RNHARK-----ORQKICQRYDQLMEAWKVKVDRIE 326  
DB 359 ERKDT---FKKDSQODIPSSKDHKEKSEFVDDITRREP---YDSVEESESSENSPVP 410  
QY 327 NNPRKAKESKTYEYKQFPE-----TRKQ----- 352  
DB 411 QRKRTSVSSSSDDYKQEDSQSGSEEDFIRKQIEMSADEADAGSDDEDFIRNOLKEI 470  
QY 353 -----REQOERFORVGAGLSATIAERS-----EHEISEIID---GLSE 389  
DB 471 SSSVESQKKEKTKGKGTACKHRLTRKSTSDDEAGRRHSWHEDEADFESPELKY 530  
QY 390 QENNEKQMRQLSVIPPMFMDAEOERVRKFINNGLMEDPMKYIKDRQFMNWTDEKEIFK 449  
DB 531 RETKSQSEELVWTG---GGGLRRFKFTIELNSTIADKYSAESQKKTSLYFDEPELEM 586  
QY 450 DKFIQHP---KNFGLIASYLERKSVDPDCVLYLYLLTKKNENYKALVRNKGKRRNQIAR 507

DB 587 ESLTSPDRSRGEGSSSLHASSFTPGTSPTSVSLDESDS--SPSHKKGESKQORAR 644  
QY 508 -----PSQEEKVEKEEDKAETKEKEEKKDEEKDEKSKENTKEKDKIDGTAE 559  
DB 645 HRPHGPLLTIEDSEEBELREELLKEQEKQKQKQKSSKSKKDKDEL--RAQ 702  
QY 560 ETEEREQATPRGRKTANSQGRKGRITRSMNNEAAAAAATAATEEPPPLPPPE--P 617  
DB 703 RRRERPKTPP-----SNLSPIEDASPTTEL--ROAAEMEBELHRSSCSEYSPSIESDPEGE 756  
QY 618 ISTE-----PVETSRW--TEEMEVAKKG----- 639  
DB 757 ISPEKIIIEQVKVYKLPATVSLYSPIDEOSIMOKESQALKSAEEMYEMMHKTHYKAF 816  
QY 640 -----LVHEGRNMAAIAKMGVTKSKAQCKNFYFNKR 671  
DB 817 PAANERDEVFEKEPLYGGLTIEDIYESLVEDYNGSVGSLLTRQEE---ENGPMQKG 873  
QY 672 RHNLDNLQOHHKQTSRKPRE--ERDVSOCEVASTV----- 706  
DB 874 REQIRLSEQIYEDPMOKITDLQKEFEYELSHSVVPOEDIVSSFIIPESHEIYDLGTM 933  
QY 707 --SAQEDF---DIEASNEE----- 729  
DB 934 VTSTEEERKLLDADAAYEELMKROOMQLTPGSSPTQAPIGEDMTSTWDFRMPDASLTS 993  
QY 730 AVKPSDESPENATSGNTEPAVEL----- 753  
DB 994 SVLSGASLTDSSTATSLIPDVKITQHFSTEEIEYVTDVTRTETQIIAIESLILTYSE 1053  
QY 754 --EPTETAPSTPSL--AVPSTKPADESVETQVNDNISAEATQMDVDQOEHSAER-- 807  
DB 1054 PSESATSVPPSDTSLTSSVSVCTDSSPITTL-DSITTYTEPVDMTKFEDESEIS 1112  
QY 808 -----GVCVCPPTAKADSVDEVVRPENHASKVEGNTKERDLDRAAS---EKVEPRDE 858  
DB 1113 SSTVFPQGIIDVP--BEISVSLDRTAPPD--GRASADHIVISLSDMASSIIESVVPKPE 1167  
QY 859 -----DLVVAQO-----INAOPEPQSDNDSATCSADEVDG 891  
DB 1168 GPVADTVSTDLILLISEKDPVKKAKKGTNGIILEAYRDKKELEAERTKSLSETVEDH 1227  
QY 892 EPERQRMFPM-----DSKPSLLNPTGSILVSS--PLKNPLDLPOLQHRA 934  
DB 1228 PPSVIALPMKEQLSTTYFTSGEFGQEKPAQSLPGSPSVSSLPKPRPF----- 1278  
QY 935 AVIPPMVSCPCNIPIGIPVPSGYALYORHIKAMHESALLEQRQREQOIDECSRSTSPC 994  
DB 1279 -----PPPPPP-----FRSSSLDISAQPPPPP- 1294  
QY 995 GTSKSPNREWEVLOPAPHQLITNLPEGVRLPTTRTPRPPPLIPSSKTTVASEKPSFTMG 1054  
DB 1295 -----PPPPPP-----PPPPPPPPPPPPPPPT-----SPKPTILPK 1325  
QY 1055 GSIQGTPTGTYLTSHNOASYTOET-----PKPSVGSISLGL- 1090  
DB 1326 KKLIVASPVTTATPLEDAVTTLETAVLRNSGLPVTRICTTAPPVPPKPS--SIPSLV 1383  
QY 1091 -----PROQESAKSATPLYTKQBEFSPRSONSQPEGLLVRAQHEGVVVRGAGAIQEGSI 1144  
DB 1384 FTHRPEPSKPPAPKVPVLPQPTTQKPTDHPKPTGLSLTSSM-----TLNLVTSADY 1437  
QY 1145 TRGPTTKISV---ESIPSLRGSITOG--TPALP--OTGIPTREALVKGISIRMPEDSSPE 1198  
DB 1438 KLPSTPLSPHSHKNSPRFSKSLTETVVTLPSEPTPTDSSASQAITSWPL--GSPS 1495  
QY 1199 KGREAAASKGHIYEGKSGHILSDNIKNAERGRSPRTAHEISLKRSYESVEGNIKQGM 1258  
DB 1496 KDLVSVPEVPSVW-----PPVTAVEIPISSPQTYIYISGALQTF 1533  
QY 1259 SMRESVPSAPLEGLICRALPRGSPHSLKERTVLSGSIQCTPRATTESFEDGLKYPKQI 1318

```
Db 1534 S--APPVTPAP-----SSFOAAPTSVTOFLTTEVSKTEVSATRSAPSV--GL-----SSI 1579
Qy 1319 KRESPIRAFEGAITKGPYDGIITIKEMGR-----SHEIPRODILTQESRKTEPV 1370
Db 1580 SITIPPELALDNIHLKP-----QYKEDGKLQLVGDVIDLRTVPKVEKTTD--KCIDL 1632
Qy 1371 VOSTRPI---TEGSIQOTPIKFDNNSQSAIKHNKSLITGPKSLKSGMPLEIVPENI 1427
Db 1633 SASTMDVKRQITANEVYKQI-----SAVQPSIINL-----SVTSSIVTPVSLATETV 1680
Qy 1428 KVERGKYEDYKAG--EIVRSRHTSVVSSGSPVLRSTLHEAPKAQLSPGIYDDTSARTP 1485
Db 1681 TFWICTASASTTGTESLVGAHAMTT---PLQLTSKHAEPYPYKIPSDQVFPFAREAP 1737
Qy 1486 VSYQNTMSRGSP-----MNMRTSDVTIPNKNSTNHERKSTL-----TPTQRE 1527
Db 1738 IN----LSLGTPAHAVTIAITKPVTPVGVNWTGTDSTVSGQITDGEVVDLSTTKSHRT 1793
Qy 1528 SIPAKSPVGV--DPVSHSPFDPHHRGSTAGE-----VYWSHLPTQLDOPAMPFHR-- 1576
Db 1794 VVTWDESTSSVMTKIEDEKPDV-----LTAGRAVCCDVVY-----KLPFGRSC 1838
Qy 1577 -ALOPAAAA-----YLFQR-----QLSPTPGYPQYQLYAMENR 1610
Db 1839 TAQPATTLPEDRGYRDDHYQDRSGPYGYRGIGMKPMSMDNLAEAGHFFYKSKNA- 1897
Qy 1611 OTILNDYITSQOMVNL---RPDVARGLSPREQLGLPYPATRGIIIDLTMNPPTILVPHP 1667
Db 1898 ----FDYSEGTDVADLTSGRVTGEVMDYSKTTG-PYPETROVIS-----G 1940
Qy 1668 GGTSTPPMDRITYIGTOITPPRPYNSASMS-PGHPHPLAAASAEARERERERERER 1726
Db 1941 AGISTP-----QYSTARMTPPPGPYCGVSV-----1966
Qy 1727 ERIAASDLYLRGSEQPG-----RPGS--HGVYRSPSPSVRTOETML-----QQ 1770
Db 1967 --LRSSNGVYSSVATPPTSTAITOPGSIFTSTVRDLG-GIHTADAVTSLPAMHSHQ 2023
Qy 1771 RPSVFOGTNGTS---VITPLDPTAQLIMPLPAGGPGSISQGLPASRYNTAADALAALV- 1825
Db 2024 MPRSFIITGASETDIATVIGDISASLTITMESLTATIDSVP--LTTASEVPEPVG 2081
Qy 1826 --DAAASAPQ-----MDVSKTKE-----SKHEAARL-----1849
Db 2082 DESALLIIVPEEDKQOQDLDERELLEKEIKOORFAEELEWEROEIQRFREQETMVKK 2141
Qy 1850 -----ENLRSRSNAVSEQOOLEQKT-LEVEKRSVQCLY 1882
Db 2142 LEEQSMKQHLLFOQEEERQAQFMMROETLAQOQLQLEQIQOQLQQLHQLEEQKIRQIY 2201
Qy 1883 TSSAPFPGSKPOPH-----SSVYSEAGKDKG-----1908
Db 2202 QYNVDPGTSAPQITTEQAILEGQYAALEGSQFWATEDATTTASAVVAIEIPQSGWTV 2261
Qy 1909 -----PP-----PKSRYEEELTRGKTTITAAANFID---1934
Db 2262 QSDGVTVQIAPPGILLSTVSEIPLTDVVVKEEKQPKR--SSGAKVRGQYDDMGENTDDPR 2320
Qy 1935 ----VITRQIASDKDAREG--SQSSDSSSLSHRYETSDALEVISPASSAPPOEK 1988
Db 2321 SFKIVDSGVTDDEDATDRSYVSRRRRTTKKSVDTSVQTDDEDDODWDMPTRSRR--KAR 2378
Qy 1989 LQYQPEVVKANAENDPTROYEGLPHHYRQOESPSPQOQLPPSSQAGMGQVPTURL 2048
Db 2379 VKYGDSTMEADKYK-----PL-----SKVSSIAVQTVAEISVQTEPVGTI--RTFSI 2424
Qy 2049 ITLADHICQIITODFARN-----QVSSQTPQPOPTSTFQNSALYSTVPVTKTS 2098
Db 2425 RARVDKAVEIKHISAPEKTYKGLGCGQTEADSDTQSPQVLSATSPPKDKRP--TPLE 2482
Qy 2099 NRYSPESAQSVHQRGSRVSPENLVDKSRGSRPGKSPERSHVSSEPYE-----PISP- 2152
Db 2483 IGYSSHLRADSTVOLAPSPPKSVLYSPISPLSPGSALESFV---PYEKPLPDDISPO 2539
```

```
Qy 2153 -----PQVPVVEHKQDLSLLLSQGAEPABQRNDARSPGISYLPSPFFTKLENTSPMV 2205
Db 2540 KVLHPDMAKVPASPPTAKMKQMSDPKPLSPADESSRAPFQYEGYTTKGSQTM---2596
Qy 2206 KSKQOEIRFKNLNSGGSDMAAAQ-----GTEIFNLPAVTTSGSVSSRGHSPAD 2256
Db 2597 -----TSSGAQKVKRKLTPNPPPEISTGTQ-----STFSTMGTVSRRICRTN 2640
Qy 2257 PASNLG-LEDIIRKALMGSDKVEDHGVMSQPVVPG-----TANTSVVTSGETRE 2310
Db 2641 TWARAKILQDIDREL-----DLVERESAKLKKQAELEDEEKEIDAKLYLEMGINRRK 2694
Qy 2311 EGDPSPHSGGVCKPKLISKNSRKSPIPG-----QGYLGTERPSSVSSVHSEGD 2361
Db 2695 EA-----LLKEREKREAVLQGVAEQDRDYMDSDEVSTRPTRIESQHG---2737
Qy 2362 YHRQTPGWAWEDRPSSTGSTQF-----PYNPLTMRMLSSPTPTPIACAP 2405
Db 2738 -----IERPTAPQTEFSQFIPPTQTESQVLPVPTSPYTOYQYSS-PALPTQ-AP 2785
Qy 2406 SAVNOAAPHOQNRIRWEREPAPLLSAQYET 2434
Db 2786 TSYTQOSHFEQOTLYHQOVSP-----YQT 2809
```

## RESULT 16

Q9UQ35 PRELIMINARY; PRT: 2752 AA.

AC Q9UQ35;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE RNA BINDING PROTEIN  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ohtaki S., Umeke K., Sawada Y.;  
RT "Homo sapiens mRNA for RNA binding protein, complete cds.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB016092; BAA83718.1; -  
DR InterPro: IPR002965; -  
DR PRINTS: PR01217; PRICHEXTENS.  
SQ SEQUENCE 2752 AA; 109C64F181097123 CRC64;

Query Match 3.7%; Score 473; DB 4; Length 2752;

Best Local Similarity 19.2%; Pred. No. 6e-17;

Matches 554; Conservative 316; Mismatches 1121; Indels 890; Gaps 122;

Qy 33 RHQQFAVPDYRSHLEVSAQSLQOQQOQQOQLR-----RRPSLLSEFHGS 79  
Db 247 RSRSTTPAKSRRAHRSTADSASSDTSRSRSRAAAKTHTTALAGRSPS-----PAS 300  
Qy 80 DRPQERRTSYEPF-HPGSPVDHDS---LESKRPRLEQVSDSHFORVSAAVLPLVHPLPE 135  
Db 301 GR---RGEGDAPFSPGTTTQRPSPSPETATKQPSPEYEDKDKKKEKSATRP--SPSPE 355  
Qy 136 GLRASADAKKQDA-----FGGKHEAPSPISQPCGDDQDNASPSK-----LSKEEL 181  
Db 356 ---RSSTGPEPPATPPLLAERHGGSPQPLATPTLSQEPVNPPEASPTDRSPPKSEKL 412  
Qy 182 IOSMDVRDREIAKVEQQILKLKKKQOQLEEEAAKPEPEKVPSPPPVEQKHRSIVQIYD 241  
Db 413 PQSS---SSSSPPSPQTKVSR-----HASSSPSPKPA---PAPGSHREISSPTS 459  
Qy 242 ENRK---KAEAAHKIFELGLPKVELPLYNQPSDTKVYHENIKTNQVMRKLLILFFKRN 297  
Db 460 KNRSHGRAKRDKSH-----SHTPSRRMGRSRSPATAKGRSRSTP-----TKRG 504





```
Db 2436 SOSLLPPAQDQPRSPVPSAFSDQSKCLIAQTPPVAGSOSLSSGAVATT--TSSAGDH--- 2490
QY 2058 IITQDFARNOVSSOTP-----QQPTSTFQNSPSALVS-TPVTRKTSNRYSPESQAQS 2109
Db 2491 -----NGMLSVPAPGVPHSDVGEPPASTGAQPSALAAIQPAKERRSSSSSSSS----- 2539
QY 2110 VHQRPGRSRVSPENLVDRSGSRPGCKSPERSHVSSEPYEP-----ISPPQVP--VVHE 2160
Db 2540 -----SS 2592
QY 2161 KODSLLLLLSQCAEPAEORNDARSQGISYLPSPFTKLENTSPMVKSKQEIFRKLNSG 2220
Db 2593 -----GRPPEPTAKRRRRSS 2643
QY 2221 GGDSDMAAQAQCTEINFPLPAVTTSS--GSVSGRSHGFADPASNLGLEDIIRKALMGSPDD 2277
Db 2644 SSSSSPSFAKPGQALPKPASPCKPPERRSRSP-----RKPI-----D 2683
QY 2278 KVEDHGVVMSQPMGVVPGTANTSVVTSGETRREEDGDPSPHGGVCKPKLISKSNSRKS 2337
Db 2684 SLRDSRLSYSPV-----ERRRPSQPSPRDQ-----SSSSSERSGR- 2720
QY 2338 PIPGOGYLGTRPSSSVSHSEGDYHRTPGWANEDEDRPSSTGSGTQFPYNPLTMRMLSTP 2397
Db 2721 -----RGQRGDSRSPS-----HKR-----RRET-----SPRPMHRRSSRS 2751
QY 2398 P 2398
Db 2752 P 2752

RESULT 17
ID Q9FND5 PRELIMINARY; PRT; 2910 AA.
AC Q9FND5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SIMILARITY TO HEAT SHOCK PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned pl clones.";
RL DNA Res. 4:291-300(1997).
DR EMBL; AB006702; BAB11602.1; -.
DR EMBL; AB017062; BAB11602.1; JOINED.
KW Heat shock.
SQ SEQUENCE 2910 AA; 375351 MW; A847EC3FE1427DF7 CRC64;
```

```
Query Match 3.7%; Score 467.5; DB 10; Length 2910;
Best Local Similarity 19.5%; Pred. No. 1.3e-16;
Matches 495; Conservative 366; Mismatches 995; Indels 683; Gaps 122;

QY 144 KKDPAFGCKHEAPSPISGQCGDDONASPSKLSKEELIQSDRV--DRETAKEVQQLK 201
Db 133 KKEEDTTGKPEVS--VEKPVIEEDQTEAKHSLQEEDIGNISKVLDTTPVKVDEYDIE 190
QY 202 LKKKQQLLEEAAPPEKPVPPV-----EOKHRSTVQ-----II 239
Db 191 -KSLNSVCEEIPITKDEVRETDRTVTSYNGTAEAHNATVSVEEISRNGDNVTNVS 249
QY 240 YDENRKAAEAHKEIFGLGPKVELPLYNQ-PSDTKVYHENIKTNQVMRKKLILFFKRRNH 298
```

```
Db 250 EDQATDCEPLHDV-ETIKREAE-PFYKTVVEDAKIV--NTEETTAHESKIL---KEDNH 302
QY 299 AR-----KQEKQIKQICQVDQLMQAEKVKVDRIENPRRKA 333
Db 303 QGEYAESYEATKNSDAEQSSREVTVDKKEEDITIONEEVQES-----PSYMESPTIQG 357
QY 334 K--ESKTRREYKQPEIRKQKQEQORFORVQRGAGLSATIARSEHEITSEIIDGLSEQ 391
Db 358 EDIESKASLDHEEMDKITKDTQEQEHV-----LVRDVPVPQSETLVTEAKTAET--- 407
QY 392 NNEKOMRQLSVPPMFMFAEQRRVKFINM-----GLMEDPMKVYKDRQFMVMTDH 443
Db 408 -----FSVQEAELKTNINESEAHSAIGGEDQGETKEN-----TEP 444
QY 444 EKEIFKDFIQHPNFGLIASYLE-RKSVPCVLYYYLTKKNENKALVRRNYGRRGRN 502
Db 445 SKDLKDDKEQEDSETVKTIISSDEVRSDDVQAEV-----GEHTEPCSSSEIKDDRHRGD 498
QY 503 QOIARPSQ-----EKVEEKED-----KAEKTEKKEEKQDE-----EKDE 540
Db 499 ESIEVKAKETGPEITAVDGSVHEIETTERVLLAEKEDKEEIKIDEEPSLNAIERAE 558
QY 541 KEDSKENTKDKIDG---TAEETEERQATPRGRKTANSQGRKRGRITRSMNTNEAAS 597
Db 559 TENKIVIEEPEIVNNEETSVHSESLSKENAEVPEAVKNSDQTE--QISREVTVD----- 611
QY 598 AAAAAATEEPPPLPPPEPISTEPVETSRWTEEMEVAKKGLVHEGRNWAIAKMGVK 657
Db 612 ----RAKEE---DIAPKTEIEQERPSEKASLEPEEV----- 642
QY 658 SEAOCKNFYFKRRHNLNLLQOHKQKTSRKPRERDVSOCEVASTVSQAQEDDEIAS 717
Db 643 -----DHISNTEEEHVL-----ERDVQOCETIESE-AVETKEDTQPS 680
QY 718 NE-EENPEDSEVAVKPSEDSPENATSGNTEPAVELEPTTE-TAPSTSPSLAVSTKPA 775
Db 681 LDLKEDKETEAEATFKTVFSSDEVRS-----AVOEQFGEHTPECSS--EIKDESHG 731
QY 776 EDESVEVQVNDNISAEAEQMDVQOQHSAEQSVCDPPATKADSDVDEVRVPPENHASK 835
Db 732 KEESVEVKSQETVQDENTE-----DKHDV-LKVPSTESK 765
QY 836 VEGDNTKERDL---DRASEKVEPRDELDLWAOQINAOPEPQSDNDSSATCSADBDVGE 892
Db 766 YQG-NETPVLVSTGSEKSEKSPDLV---LNVDEEL---NDEKINV---DQVDT 814
QY 893 PERQRMFPMDSK-----PSLLNPTGSIILVSSPLKPNPLDLQLOHRAAVIPPMVST 944
Db 815 QIMEEPIGLDSNGAEAEQIDQNTITEILVAKPV--SLLDVKSVEQMQK--PKLES-- 868
QY 945 PCNIPIGT-----PVSGYALYOR-HIKAMHESALLEEQROEQDLECRSSTS 992
Db 869 PSEVSEETSKTVDEKIEKPEEVTLYQEGVDGSGYLETKEETVSVPESEIE----- 922
QY 993 PCGTSKSPNREWEVLQAPAPHOLITNLPGLVRLPTRP--RPPPLIPSSKTT--VASEK 1048
Db 923 -----EQQEERSVIDPTPLQ-----KPLESPSEVLEESSKTVDEKIEK 963
QY 1049 PSFIMGSGISQGTPTGYTLTSHNQASYTOETPKPSVGSISLGLPQOESAKSATLPYIKOE 1108
Db 964 TDSIELGEIAQ-----EERSVTDLTLPQESS---QPNEQKE-----TKLE 1002
QY 1109 EFSRPSQNSQEGLLVRAQHEGVVGTAGATQESITRGFTPTSKISVESIPSLRGSIQ 1168
Db 1003 KHEPTNEEVKSDEVI-----EVLASPSKELEGE-----TVVAENIENIKENEQ 1049
QY 1169 TPALPQTCIPTEALVKSGISRMPIEDSSPEKREAAKSHVIVBEGKSHILSYDNIKNA 1228
Db 1050 AAEKIQKSLVTVQVESPSLL---FSSEODHVTVAEE---IVDEKAKEEVPMLQIKNE 1103
QY 1229 REGTRSPRTAHEISLKRSYESVEGNIKQGMGMRSPVSAPLEGLICRALPRGSPSHDLKE 1288
```



QY 235 -----IVQIIYDE-----NRKAAEAKHIEGGLPKVELPLYNQ 269  
Db 1361 PSETRDISIEEIEKESQKVKTSKDSQAQFPSPKKEHENPELVDDLSR-----1411  
QY 270 SDTKVYHENIKTNVMRKKLILFFKRR-----NHARKOREQKICQYDQJLMAWEKKVDRI 325  
Db 1412 ---RASYSDESSESENSESVARRRRRTSIGSSSEYKQEDSQSGDEDFIRKQI--I 1466  
QY 336 ENNPRKAKESKTRYEYKQPEIRKQREOQERQOR---VQORGAGLSATIARS-----376  
Db 1467 EMSADEASGSEDEEFIRSQKELIGGVYTESQKRETKGKSPAGKRRRLTRKSTSPDD 1526  
QY 377 -----EHEISEITIDLSE---QENNEKQMRQLSVIPPMFMFAEQRAQVRFINNGLME 425  
Db 1527 DAGRRHSHWDEDETFDESPELKPRETKSQSEELVAVG-----GGGLRRFKTIELNSVT 1582  
QY 426 DPMKVYKQRMVWTDHEKEIFDKFTQHPK-----NFGLIASYLERKSVDPCLVY 478  
Db 1583 DKYSAESSQKKTLYFDEPELEMESLTDSPEDRSRGEGSSSLHASSFTPGTSPTSV--S 1640  
QY 479 YLTKKNE---NYKALVRNRYKRRGRNQIARPSQOEKVEEKEEDKAEKTEKKEEKKDE 535  
Db 1641 SLDESDSPSHKKGESKQKARRHSHGLPLPTIEDSEEEELKEEELKEQKQREL 1700  
QY 536 BEKDEKESKENTREKIDGTABETEEREQATPRGRKTANSQGRKGRITRSMTNEAAA 595  
Db 1701 EQQKSSKSSKKDKDEL--RAQRRRRPKTPP-----SNLSPIEDASPTTEL-RQAAE 1752  
QY 596 ASAAAAATEPPPLPPPE--PISTE-----PVETSRW--TEEMEVAKKG--639  
Db 1753 MEELHRSSCSEYSPSIESDPDEGFEISPEKIIEVOKVYKLPTAVSLYSPTEBQSVMKQEGA 1812  
QY 640 -----ASTVSAQED-----DIASNEE-----720  
Db 1930 POEDIVSSYIIPESHEIVDLGSMVTSTSEKLLDADAAYEELMKRQOMQVTDGSSLIQ 1989  
QY 721 -----ENPEDSEVAVKPSDESPENATSRGNT-PAVEL-----753  
Db 1990 TTMGDDMAESTLDFRVQDASLTSSILSGASLTOSTSATLSIPDVKITQHFSTEFEDE 2049  
QY 754 -----EPTTETAPSTPSLAVP-STKPADESESVETQND 786  
Db 2050 YVDTYREIQEIIAHESLILYSEPSATSVPPSDTPSLTSSISSVCTTDSPPVTILD 2109  
QY 787 SISAEABQMDVQOESABE-----GSVCDPPP-----815  
Db 2110 SLTTYTEPADVITKFKDSESSISSTYFPGSIDYFEDIGVSLDRITIPESRTNADQIMIS 2169  
QY 816 -----ATK-----ADSDVDVEVRVPENH---ASKVEGNTKERDLDRASEKVEP 855  
Db 2170 PPGTAPSTESVATKPERPQADTISTDLPISEKELIKGKETGOGIILEVDJAYDKREE 2229  
QY 856 RDEDLV---VAQQINAORPEQSDNDSSATCSADEDVGEPERQRMFPMDSKPSLLNPTG 912  
Db 2230 SEAEITKISLPETGLATPPSQTKPEQSGPHSVSGEISGQ-----EKPYRSPG 2279  
QY 913 SILVSS-PLKNP-----IDL-----PQLQRAAVIPPMVMTCPNIP 949  
Db 2280 GLPVSTHPSKSHPPFRSSSLDISAQPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 2339  
QY 950 -----IGTPVSGYALYQRHIKAMHESALLEQRQREQIDLECRSSYSPCSTKSPNREW 1004  
Db 2340 KRKLAAAPVAPTAVTAHAADAFTVEATAARRNG-----LPATKICAAAPPP-----2388

QY 1005 EVLOPAPHOLITNLNPEGVRLPTTRPTRPPLIP-----SSKTTVASBKPSFIMGSSISO 1059  
Db 2389 --VPPKPSIITGLVFTHRPEASKPPIAPKFAVPEIPVTQKTTDTCTPKPT-----2437  
QY 1060 GTPGTLYLFSHN---QASYTOETPKPSVSGISLGLPROQESAKSATLPYIKQBEFSPRSON 1116  
Db 2438 GLPLTSSNLSNLVTSADY--KLPSPT-----SPLSPHSNKS-----SPRYSK 2477  
QY 1117 SQPEGLLV--RAQHEGVVRGTAGALQEOGSIYRGTPTSK-ISVESIPSLRGSTQGTQTPALP 1173  
Db 2478 SLMETYVITLIPSEPGTPTDSSAAQAITSWPLGSPKDLVSLVETFSV-----VPPMT 2530  
QY 1174 QTGIPTEA-----LVKGISIR-MPIEDSSPEKREAAASK 1207  
Db 2531 STEIPSASQPTLYTSGALGTFSTVPATVATSLFQVPTSLTQFLPAEASKPEVASVSSAVP 2590  
QY 1208 GHVYEGSGHILSYDNINKNAREGTRSPRTAHEISLKSYESVEGNIKOGMSMRESPVSA 1267  
Db 2591 -----SVAPRSVSIPIPPPEALDRHQYKENGKL-----2619  
QY 1268 PLEG--LICRALPRGSPHSDLK--ERTV-LSGSIHQGTPTATTESFEDGLKYPKQIKRES 1322  
Db 2620 PLIGDAIDLRTIPK---SEVKVTEKMDLSASAMDVKRQJTANEV-----YRROI SAVQ 2670  
QY 1323 PPIRAFEGAITKGKPYDGIITTIKEMGRSIIHEITPRODIILOESRKTPEVVSQSTRIIEGSI 1382  
Db 2671 PSILNSAASSLIGTP-----VTMDS-KTIVAVTCT-----2699  
QY 1383 SQGTPIKFDNNGSGSAIKHNKVSILITGPKSLSRGMPPILEIVPENIKVVERGKYEDVKAGE 1442  
Db 2700 --DTIYVTGTESQVIGIEHAVTS-----PLQL-----2724  
QY 1443 TVSRHTSVVSSGSPVSLRSLTHEAPKAQLSPCIYDDTSARRTPVSYQNTMSRSGSPMMNRT 1502  
Db 2725 --TTSKHTEL-----QYRKPSSQAFPMIRDE-----APINLSLGPSTQAVTLAVT 2767  
QY 1503 SDVTIPPNKSNHHERKSTLTP--TORESI--PAKS--PVPGVDPDVVSH-----S 1545  
Db 2768 KPVIYVPPGVINGWTDTSISQIGITDGEVVDLSTKSHRTVTMDSTSNVYTKIIEDEEK 2827  
QY 1546 PFDPHHRGSTAGE-----VYWSHLPTQLDPAMPFHR---ALDPAAAA-----1584  
Db 2828 PVD-----LTAGRAVACDMVY-----KLPPGRCSTAQOQPAATTLPELDFRGYRDDH 2872  
QY 1585 YLFOR-----QLSPTPGYPQSOXYOLYAMENOTIILNDYITSQOMQNL---1627  
Db 2873 YOYDRSGBYRGIGGKMPKPSMDSNLAEGHFFYKSKNA-----FDYSGGTEAAVVDLTSFG 2927  
QY 1628 RPDVARGLSPREQPLGLPYPATRGIIDLNNPPTILVPHPGGTSTPPMDRITYIPGTQIT 1687  
Db 2928 RVSTGEVMDYSSKITG-PYPETROVISGV-----GISTPOYST-----ARMT 2968  
QY 1688 FPPRP--YNSASMSGPHPLAAASAEERERERERERERERERERERERERERERERER 1746  
Db 2969 PPGPGYQVGVSV-----LRSNGVYVSSVATPIPS 2998  
QY 1747 -----RPGS--HGYYRSPS-----PSVTOETMLQORPS-----VFOGTNGTSV-ITPL 1787  
Db 2999 TFAITQPGSIFSTVTRDLSGIHTDALTSLSHQSOQMPRSPYFITTGASETDISVTSI 3058  
QY 1788 DPTAQLRIMPLPAGGPSISQGLPASRYN-----TAADALAALV-----DAAAASA 1831  
Db 3059 DINASLQIITM-----ETLPAETMDSVPLTITASEVFEVGVVEESTLLIVPDEKQ 3110  
QY 1832 POMDVSK-----TK 1840  
Db 3111 QOOLDERELLELEKIKOORFAEELEWEREQEIQRFREQEIMVOKKLEELQSMKQHLLYQ 3170  
QY 1841 ESKHEAARL---EENLRSASAAVSEQQOLEQKT-LEVERKSVQCLYTSFAFSGKPOPH--1895  
Db 3171 EERQAOQMMRQETLAAQOQLQLEQIQOOLHQOLEQKRLQIQYNYEYPSGTASPOTT 3230  
QY 1896 -----SSVYSEAGKDKGPPKPSRYEEELRGT--1925

Db 3231 TEAILEGQYVATESQFOWATEDATTASTVAIEI-----PQSQGWYVQSDGVTOY 3283  
Qy 1926 -----TIAANFIDVILITROASDKDAREGSSOSSSSLSHRYETPSDAIEVIS 1977  
Db 3284 IAPGILSTVSEIPLDVTVVKEEKPKK--RSSGAKRGQYDEMGESMADPRNLKLIYD 3341  
Qy 1978 PASSPAPPOEKLQTY-----QPEVVKANAENDPTROYEGPLHHVRPQO-----ESP 2024  
Db 3342 SGVQTDDEETADRIYASRRRTKSVDTSVQTDQDQDEWDPMSRRKARTGYGSDTA 3401  
Qy 2025 SPQOOLPPSS-----QAEGMGQVPRTHRLITLADHICQII-----TODFA 2064  
Db 3402 EGDTKPKPSKVSSVAVQTVABEISVQTEPLGTI-RTPSIRARVDAKVEIKHISAPEKTYK 3460  
Qy 2065 RNOYSSOTPOQPTSTFONSPSALVSTPVRTKTSNRKSPESQAOQSVHHOR-----PG 2116  
Db 3461 GSGLCQETETPDPT-----QSPPYMGATSPKDK--KRPTPLEIGYSSSHLRADPTVQLAPS 3515  
Qy 2117 SRVSPENLVKSRGSRPGKSPERSHVSSEPYE-----PISP-----POVPVVHEKQD 2163  
Db 3516 PKSPKVLVSTISPLSPGHLEPAFV---PYEKPLPDDISFQKVLHPDMAKVPASPKTA 3572  
Qy 2164 SULLLSQGAEPABORNDARSPGISYLPSPFTTK-LENTSPMVKSKKOEIPKLNSSGGG 2222  
Db 3573 KMQORSMDPKPLSPADESSRAPFOYSEGFTAKGSQTSQTKVKRTL-----P 3623  
Qy 2223 DSDMAAQPTEIFNLPAVTTVSGSVSSRGHGFADPASNLG-LEDIIRKALMGSDDKVED 2281  
Db 3624 NPPPEASTGTQ-----STYSTMGTASRRMCRMTNMAKALQDIDREL-----DLVER 3673  
Qy 2282 HGVNMQPMGVVPG-----TANTSVTSGETRRREGDPSPHSGGVCKPKLISKNSRKS 2336  
Db 3674 ESAKLKQAELEDEEKEIDAKLYLENGINRKEA-----LLKEREKR 3719  
Qy 2337 SPIQ-----QGYLGTERPSVSVHSEGDYHROTPCGAWEDRPSSTGTQF----- 2383  
Db 3720 AYLAQVAERDYMDSSESVSTRPSVESQHG-----IERPTAPQTEFSOFI 3766  
Qy 2384 -----PYNPLTMRMLSTPTPIACAPSAVNOAAPHQONRWEREPAPLLSAQ 2431  
Db 3767 PPQQTQEAQLVPTSPYTOYQYSS--PALPTQ-APTPTQQSHFQOQTLYHOQVSP----- 3819  
Qy 2432 YET 2434  
Db 3820 YQT 3822

## RESULT 19

Q9QYX7  
ID O9QYX7 PRELIMINARY; PRT; 5038 AA.  
AC O9QYX7  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE ACZONIN.  
GN ACZ.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=99439764; PubMed=10508862;  
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,  
RA Kilimann M.W.;  
RA "Accozinn, a 550-kd putative scaffolding protein of presynaptic active  
RT zones, shares homology regions with rim and bassoon and binds  
RT profilin.";  
RL J. Cell Biol. 147:151-162(1999).  
DR EMBL; Y19185; CAB60731.1; -.  
DR HSSP; P04410; IA25.

DR MGD; MGI:1349390; Acz.  
DR InterPro; IPR000008; -.  
DR InterPro; IPR001427; -.  
DR InterPro; IPR001478; -.  
DR Pfam; PF001168; C2; 2.  
DR Pfam; PF00595; PD2; 1.  
DR ProDom; PD000535; -; 1.  
DR ProSITE; PS00499; C2\_DOMAIN\_1; UNKNOWN\_1.  
DR ProSITE; PS00004; C2\_DOMAIN\_2; 2.  
DR SMART; SM00239; C2; 1.  
SQ SEQUENCE 5038 AA; 547541 MW; 577A7B2530F54C5C CRC64;

## Query Match

3.7%; Score 467.5; DB 11; Length 5038;

Best Local Similarity 18.3%; Pred. No. 2.7e-16;  
Matches 562; Conservative 393; Mismatches 1079; Indels 1029; Gaps 133;

Qy 21 PPHSVQYTFNTRHQQFAVP-----DYRSHLEVSAQSL-----LQOQOQOQLR 66  
Db 1140 PLEEKPIPDQQLPPDAKPSASEGEKRLKKAHVQIPEEGPIGKVASLACEGQOPD 1199  
Qy 67 RPSLLSEFHGS-----DRPQERTSYEPFHPGSPVDHDSLEKRPRLQEQVSDSHF 119  
Db 1200 TRPEDL-----PGATPQTLPKDROKESRDVTPQAEGTAKEGRG--EPSKDRTEEKDK-- 1251  
Qy 120 QRVSAAVLPLVHPLPEGLRASADAKDPAGGKHEAPSPISGQPCGDDQNASPSKLSKE 179  
Db 1252 SDTSSSQP---KSPQGL-SDTGYSSDGISGLGEIETSLPS-----DEKDLKGLKKD 1301  
Qy 180 ELIQ-SMDVRVDREIAKVEQQLKLLKKQQQ--LEEEAAKPEPEKVPSP--PPVEQKHRS 234  
Db 1302 SFSQESSPSSDLAKLESTVLSLEAQASTLVCEAKKATQPKQ-VSPEQPDQOQKTOT 1360  
Qy 235 -----IVQIYDE-----NRKAEAEAHKIFEGIGPKVELPLNQP 269  
Db 1361 PSETRDISIEETIKESQEKVKTSKDSAQGFPSPRKEHKNELVDLSPR----- 1411  
Qy 270 SDTKVYHENTKTQNMVRKKLILFFKRR---NHAKOREKICQRYQQLMEAWKKVYDRI 325  
Db 1412 ---RASVDSVEDSSESENSPVAREKRTSIGSSSEYKQEDSOGSGEDEFIRKQI--I 1466  
Qy 326 ENNPRKAKSKTRYEYKOFFPEITRKOREQOERFOR--VGORGAGLSATIAARS----- 376  
Db 1467 EMSADEDASGSEDEEFIRSQLKEIGVTESQKREETKGGKSPAGKHRLTRKSTSFDD 1526  
Qy 377 -----EHEISLIIDGLSE---QENNEKOMRQLSVPPMMFADAEORRYKFINNGLME 425  
Db 1527 DAGRRHSHWDEDDTDFESPPELKFRTKQSESEELVWAG---GGGLRRFRFTIELNSTVT 1582  
Qy 426 DPMKVYKDRQFMNVVWTDHEKEIFKDKFIQHPK-----NFGLIASYLERKSVDPCLVY 478  
Db 1583 DKYSAESSQKKTLYFDPEPELEMSLTDSPDRSGEGSSLSHASFPTCTSPTSV--S 1640  
Qy 479 YLTAKNE---NYKALVRNRVYKGRNRQOQIARPSQEKVEKEKEDKAEKTEKKEEKKE 535  
Db 1641 SLDEDSDSPSHKKGESQKQKARHSHGHPLLPTIEDSSESEELREBELLEKQEKQEL 1700  
Qy 536 EEKDEKEDSKENTKEKDKIDGTAETEEREQATPRGRKTSQNSQGRKGRITRSMNTNAAA 595  
Db 1701 EQOQRKSSSKSKKDKDEL--RAQRRRRPKTPP-----SNLSPIEDASPTTEL-RQAAE 1752  
Qy 596 ASAAAAATEEPPLPPPPPE--PISTE-----PVETSRW--TEEMEVAKKG- 639  
Db 1753 MEELHRSSCSEYSPSIESDPGEGFEISPEKIIYQVKVYKLPATVSLYSPTDEQSVMQEGA 1812  
Qy 640 -----LYEHGRNW 647  
Db 1813 QKALKSAEMEYEMMHKPHYKAPPAANERDEVEFEKEPLYGGMILIEDIYVESLVEDYNG 1872  
Qy 648 AATAKMGVTKSEAOCKNFYNYKRRHLDNLQOQ-----HKQKTSRKPREERDVSQCESV- 702  
Db 1873 SVDSGLTRQDE---QNGFMPQRCQKIRLREQIYDDPMQKITDLQKEFEYESLSHIV 1929

QY 703 -----ASTVSAQEDB-----DIEASNEE----- 720  
Db 1930 POEDIVSSYIIPESHEIVDLGSMVTSTSEBKLLDADAAYEELMKRQMQVTDGSSLIQ 1989  
QY 721 -----ENEDSEVEAVKSEDSPENATSRGNTPE-PAVEL----- 753  
Db 1990 TTMGDDMAESTLDFRDVODASLTSSIIISGASLTOSTSATLSIPDVKITQHSTEEFEDE 2049  
QY 754 -----EPTTETAPSTPSLAAP-STKPAEDSESVETQND 786  
Db 2050 YVTDYTRIEQIEIAHESILITYSEPSSESATVPSPDPSLTSSISSVCTTSSSPVTLLD 2109  
QY 787 SISAEATQMDVDQOEHSAEB-----GSVCDPPP----- 815  
Db 2110 SLTTVVYEPADVIKFKDSEBSEISSTYFPGSVIDYEDIGVSLDRTITPESRTNADQIMIS 2169  
QY 816 -----ATK-----ADSDVDEVRVPENH-----ASKVEGDNTERDLDRASEKVEP 855  
Db 2170 FPGIAPSITESVATKPERQADTISTDLPISEKELIKGKETGDIILEVLDAYKDRREE 2229  
QY 856 RDEDLV---VAQOINAOPEQSDNDSSATCSADEVDGEPERQRMFPMDSKPSILLNPTG 912  
Db 2230 SEAEITKISLPETGLATPPSQTREQPGSPHSVSGEISGQ-----EKPTYRSPG 2279  
QY 913 SILVSS-PLKPNP-----LDL-----POLQHRAAVIPPVMSCTPCNIP 949  
Db 2280 GLPVSTHPSKSHPPFRSSSLDISAQPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 2339  
QY 950 -----IGTPVSGYALYORHIKAMHESALLEEORQORQIDLECRSSTSPCGTSKSPNREW 1004  
Db 2340 KRKLAAAPVAPTAVTAHADAIVTAVATAARRSNG-----LPATKICAAAPPP----- 2388  
QY 1005 EVLOPAPQLITNLEPEGVRPLTPTPTPPPPPLIP-----SKTTVASEKPSFIMGGSISQ 1059  
Db 2389 --VPPKSSIPTGLVTHREPEASRPPIAPPAVPEIPVTQKTDTDCPKPI----- 2437  
QY 1060 GTPQTYITSHN---QASYTOETPKPSVGSISLGLPROQESAKSATLPYIKOEESPRSON 1116  
Db 2438 GLPLTNSMLNLTADY---KLPSPT-----SPLSPHSNKS-----SPRYSK 2477  
QY 1117 SQPEGLLV--RAQHEGVVGTAGAIQESITRGTPTSK-LSVESIPSLRGSITOGTALP 1173  
Db 2478 SLMETYVITLPSPEPGPTDSSAAQAITSWPLGSPPKDLVLETVFSV-----VPPMT 2530  
QY 1174 QTGIPTEA-----LVKGSISR-MPIEDSSPEKGREEAASK 1207  
Db 2531 STEIPSASQPLTYTSGALGTFSVTPATVATSLFQVPTVSLTQFLPAEASKPEVSAVSNAP 2590  
QY 1208 GHVIEKSGHILSYDNKNAREGTRSPRTAHEISLKRSESVESVEGNIKQGMRESVPVA 1267  
Db 2591 -----SVAPRSVSIPPEPLALDRHQYKENGKL----- 2619  
QY 1268 PLEG--LICRALPRGSPHDLK--ERTV-LSGSTMQGTPTATTESFEDGLKYPKQIKRES 1322  
Db 2620 PLIGDAIDLRTIPK-----SEVVKTEKCDLMSASAMVYKRTTANEV-----YRROISAVQ 2670  
QY 1323 PPIRAFEGAITKGPYDGIITIKEMGRSIIHEIPRODILTQESRKTPVEVQSTRIIEGSI 1382  
Db 2671 PSILNLSAASLGTP-----VTMDS-KTAVAVTCT----- 2699  
QY 1383 SQGTPIKFDNNSGQSAIKHNKVSILITGPSKLSRGMPPLEIVPENIKVVERGKYEDVRAGE 1442  
Db 2700 --DTTIVTTGTESQVGHIAVTS-----PLQL----- 2724  
QY 1443 TVRSRHTSVSSVSGPSVLRLSLTHEAPKAQLSPGIYDDTSARTPVSYONTMSRGPMMNRT 1502  
Db 2725 -TTSKHTEL-----QYRKPSQAQFPFMRDE-----APINLSLGPSTQAVTLAVT 2767  
QY 1503 SDVTIPPNNKSTNHERKSTLAP--TQRESI---PAKS--PVPBGVDVYVSH-----S 1545  
Db 2768 KPVTVPPGVNNGTDTISQGITDGEVDLSTSKSHRTVVVTMDDESTSVNVVTKIIIEDEK 2827  
QY 1546 PFDPHHRGSTAGE-----VYWSHLPTQLDPAMPFHR---ALDPAAAA----- 1584

Db 2828 PVD-----LTAGRAVCCDMVY-----KLPFGRSCTAQOQPATTLPEDRFGYRDDH 2872  
QY 1585 YLFOR-----QLSPTPGYPQYOLYFAMENRTOTILNDYITISQOMQVNL--- 1627  
Db 2873 YOYDRSGPYGYRGIGMKPMSDNTLAEAGHFFYKSKNA-----FDYSGGTEAAVADLTSG 2927  
QY 1628 RPDVARGLSPREQPLGLPLYPATRGIIIDLNTNPPPTILVPHPGGTSTPPMDRITYIPGTOT 1687  
Db 2928 RVSTGEVMDYSSKTTG-PYPETROVISG-----GISTPQYST-----ARMT 2968  
QY 1688 FPPRP--YNSASMSPGHPTHLLAAASAEERERERERERERERERERERERERERERER 1746  
Db 2969 PPGPGQYGVGSV-----LSSNGVYSSVATPIPS 2998  
QY 1747 -----RPGS--HCYVRSRPS-----PSVRTQETMLQORPS-----VFQGTNGTSV-ITPL 1787  
Db 2999 TFAITTPGSIPTVTRDLSGIHTDAITSLALHQSPMRPSYFITTGASSETDISVTSI 3058  
QY 1788 DPTAQLRIMPLPAGGPSISQGLPASRYN-----TAADALAALV-----DAAASA 1831  
Db 3059 DINASLOTIIM-----ETLPAETMDSVPTLTASEVFSEVVEESTLLIIVDEDKQ 3110  
QY 1832 PQMDVSK-----TK 1840  
Db 3111 QQLDLERELLELEKIKOORFAEELEWERQETORFREOEKIMVQKKLEELQSMKQHLLYQ 3170  
QY 1841 ESKHEAARL--EENLRSRAVSEQQOLEOKT-LEVEKRSVQCLYTSSAFSGKPOPH-- 1895  
Db 3171 EERQAOFMKROETLAQOOLQOLEQIQOOLHQLEQKLRQIQVYNYEPSTASPOTT 3230  
QY 1896 -----SSVYSEAGKDKGPPPKPSRYEEELRTRGKT-- 1925  
Db 3231 TEQALLEGQYVATGSGQFOWATEDATTASTVAIEI-----PQSGQWYTVQSDGVTOY 3283  
QY 1926 -----TITANFIDVITITROIASDKDAREGSSQSDSSSSLSHSHRYEFTPSDAIEVIS 1977  
Db 3284 IAPPGILSTVSEIPLTDVYVKEEKQPKK--RSSGAKVRGOYDEMGSMAADDPRNLKKIVD 3341  
QY 1978 PASSPAPPOEKLOTY-----QPEVVKANQANDPTROYEGPLHHVYRPOQ-----ESP 2024  
Db 3342 SGVOTDDEETADRYASRRRTKSKSVDTSVOTDDEQDMDPMSRRKARTKGYGDSTA 3401  
QY 2025 SPQOQLPPSS-----QAEQMGQVPRTHLITLADHICQII-----TQDFA 2064  
Db 3402 EGDKTTPSKSVSSVAVQVVAEISVQTEPLGTI-RTPSIRARVDAKVEIILKHSAPETYK 3460  
QY 2065 RNQVSSOTPOQPTSTFQNSPSALVSTPVRTKTSNRYSPEQAQSVHHQR-----PG 2116  
Db 3461 GGSIGCQTEPTDPT--QSPPYMGATSPPKDK--KRPTPLEIGYSSSHLRADPTVQLAPS 3515  
QY 2117 SRVSPENLVKSRGRSPKSPERSHVSSEPYE---PISP-----POVPVVEHQD 2163  
Db 3516 PPKSPKVLVISPISLPGHALEPAFV---PYEKLPPDDISPOKVLHPDMAKVPASPRTA 3572  
QY 2164 SLLLSQRGBAPQRBQNDARSPTSISYLPSEFTK-LENTSPMVSKKOEIERKLNSSGGG 2222  
Db 3573 KMQORSMDPKPLSPTADESSRAPFOYSEGFTAKGTSQTSQTKKKVETL-----P 3623  
QY 2223 DSDMAAQPQTEIFNLPAVTTSGVSSRSGHGFADPASNGL-LEDIIRKALMGSDDKVED 2281  
Db 3624 NPPPEEASTGTQ---STYSTMGTASRRRMCRTNTMARAKILQIDREL-----DLVER 3673  
QY 2282 HGVMSPQMGVVPF-----TANTSVTSGTTRREGDPSHSGGVCKPKLLSKNSRKS 2336  
Db 3674 ESAKLRKQAELEDEEKEIDAKLRYLEMGINRKEA-----LLKREKER 3719  
QY 2337 SPIPG-----QVGLGTERPSSSVSHSEGVDYHRQTPCWAMEDRPSSTGTOF----- 2383  
Db 3720 AYLOQVAEDRDYMSDSESVSTRPSRVSQHG-----IERPTAPQEFESQFI 3766  
QY 2384 -----PYNPLTMRMLSTPTPTACAPSAVNQAAPHQONRIWIEREAPLISAQ 2431



QY 966 AMHESALLEEORQR-----QEQIDLE-----CRSTSPCGTSKSPNREWEVLQ 1008  
Db 1912 SEEDGAASDDTRTRRRYAYIEDDDSDGGIKPGVHRPDTPEERQKFIQOEEL-- 1969  
QY 1009 PAPHOLITNLGVELPTRTRPPPLIPSKTIVASEKSFIMGISQGTPTGYLTS 1068  
Db 1970 --KRLMAEKNEGAKIAAT-----PLTPLKSGVTASEK-----RTPG----- 2005  
QY 1069 HNOASYTOTETPKPSVGSISLGLPROQESAKSATPLYIKOEESFSPRSQNSQPEGLLVRAQH 1128  
Db 2006 -----RAASGDSLSVTPLSVIRO---AKVLIDIDL-----QR 2034  
QY 1129 EGVRGTAGATQEGSIITRTPTSKTSVSPSLRGSITQGTALPQTGPTALYKGSIS 1188  
Db 2035 KGETIGDLDVDE-----SELDDAELPD-----DLPED-----MEDAIA 2068  
QY 1189 RMPIEDSDPEKREAAASKGHVIEGKSHILSYDNINKNAREGTSRPTAHEISLKRSYE 1248  
Db 2069 RMVEEE---EQFSAEVAAR-----ELPGAEE----- 2091  
QY 1249 SVEGNIKQGMRESVPASPLEGLICRALPGSPHSDLKERTVLGSGSIMOGTPRATTESF 1308  
Db 2092 -----VLRTPSKSQ-----TSRVMPSPASQNSAPS 2118  
QY 1309 EDGLKYPKQ-----IKRESPIRAFEAGITKGPYDGIITIKEMGRSIEHPRQDILTQES 1364  
Db 2119 TSLQEPHKKRLPMPHTMPLLRHQFPISAGPSHSPASIV----- 2158  
QY 1365 RKTPEVOSTRPTIGTSISQGTPTKFDNNSGOSAIKHNKSLITGPKSLRGMPPLEIVP 1424  
Db 2159 --PPhAAQGMHPLQRHLSQTV-----PQAMHLLQNALSA-----LQGP----- 2198  
QY 1425 ENIKVVERGKVEDKAGETVRS-RHTSVSSGSPVLRSTLHEAPKAQLSPGIYDDTSARR 1483  
Db 2199 -----LGCNRY-----GSGPNSAQHPLVMSMPSAAAAAHLMOGSAVAS-----ATARP 2242  
QY 1484 TPVSYQNTMSRSPMMNRTSDVTPPPNKSTNHERKSTLTPTQRESIPAKSPVPGVDVYVS 1543  
Db 2243 VETASGNPASDPKP-RGRKKVKTPLRDLQKQOATAAATAATSTTPGSA-----SEKYKA 2298  
QY 1544 HSPDPHRRGSTAGEVYVSHLPTQLDPAMPFRALDPAAYLAFQRLSPFTG-YPSOYQ 1602  
Db 2299 QPLFKPHEDAAPSAPASQAVITRMPSLP-----PAHG-----RNHGPPSGLYPSSAD 2347  
QY 1603 L---YAMENTROT-----LNDYITSOQMOVNLRPDVARGLSPRE---OPLGLPYPATR 1650  
Db 2348 LARFYGVANQOPIPAVPCGRSPSTSGPPHLLRPQMPGLPPHSLRTPYGPPLR 2407  
QY 1651 GIIDLTNMPPTIL-----VPHPGGTSTPPMDRITYIPGTQITFPFRPNYSASMSPP- 1700  
Db 2408 GSGPPTSTPTTNSRPAYLHGAEHGGGSPGPMGGV-FSSGP---PPARHATPHLNPY 2462  
QY 1701 -----GHPTLHAAASAERERERERERERERERERERERERERERERERERERER 1750  
Db 2463 RAPPYIGNPNY-----SPRLGAPGTCGMRPGAVDYVAGPRGYSP 2502  
QY 1751 HGYVRSRSPSVYRTOETMLQRPVSFQGTNGTNSVITPLDPTAQLRIMPL-----PAGCP 1803  
Db 2503 YGYTPPPPP-----LSTPSHAATAVSSVIVSAPHLTPTNHSVPLTTHGKTPPOQTP 2553  
QY 1804 SISQGLPASRYNTAADALVALDAASAPQMDVSKTEKSKHEAARLEENLRSRAAASVSEQ 1863  
Db 2554 TQSSGPP-----PAAAPPTITSETSHKPPPLASV---ITSKLTITLEA 2594  
QY 1864 QOLBQKLTLEKRSVQCLYTSATSPGKQPHSHSVVYSEAGKGGPPPKSYBEELRTRG 1923  
Db 2595 YPIRKSPI-----AVADVSGPAEPTRSPAP-----IAEEDSG-----SAHD-----TRA 2634  
QY 1974 KTTITANFIDVIITROIADSKDAREGSRQSSDSSSSLSHRYETPSDAIEVSPASSPA 1983  
Db 2635 PSSATGTAIV-----GFGSGSGNAQWAAH-----CTGSPT 2663

QY 1984 P---PQEKLOTQYQEVVKANQAEENDPTRQYEGPLHHYRQOESPSPQOQLPSSQAEGMG 2040  
Db 2664 PRELGQSKLELEQEQ--SKLEREQEPSKL-----ELELEQEQSKLERAQREGPG 2711  
QY 2041 QVPRTHRLITLADHLCQIITQDFARNQVSSQTPQO---PPTSTFQNS----- 2084  
Db 2712 R-----EVTRELAQEQSKLELGQEQSIPETLAQEQNTLGREQCCLQEAD 2756  
QY 2085 --PSALVSTPVRTKTSNRYSPESQAQSVHHQR-----PGSRVSPENLVDSRGRSPCK 2135  
Db 2757 TPKRVLVREQSELEGLGERSKQEQVQERSKRELEPELEPCPGREASTQELAQEQSKLELGQ 2816  
QY 2136 SPERSHVSE-----PYEPIPPQVPPVHEKQDLSLLLSORGAEPAPQORDNARS 2184  
Db 2817 ESKLELEQEQNKQELGCLQEAGTPQVLPVPSRLGLGERSKQEQSQEQSTOELE 2876  
QY 2185 PCSISYLSFFTKLE 2199  
Db 2877 PAQISPPPERVADKLE 2891  
RESULT 21  
Q9VH10 PRELIMINARY; PRT; 3111 AA.  
ID Q9VH10  
AC Q9VH10;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CG3996 PROTEIN.  
GN CG3996.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,





Db 2276 SRKS-AKSPYTS--PSR-----NR-----QSPSPTRSPERKSKSPYTSPPA 2314  
QY 1609 TRQTLINDYITISQOQVNLRLP-DVARGLSPREQPLGLVPATRGILDLTNMPTILVPH 1667  
Db 2315 RKPHNDLPISDDLEYKYRVLIVRSKSNLAKRMNDPNKPAIHLEM---ILSPP 2371  
QY 1668 GGTSTPPMDRITY---IPGTQITFP-----PR-PYNSA 1696  
Db 2372 DADAIPTTGELHRIVRVLEKLSPAKTSKRSRSRPTIEDIKRQMRDEKPKPTPVHNL 2431  
QY 1697 S---MSPGHPHAAAAAERERERE-----KERERERIAAASDLVLR 1739  
Db 2432 ERIVSSGPGRPEPTAELEERIRILEQEHKFDKTKQDYKAFNKLKDVISPLSDFDEF 2491  
QY 1740 PGSEOPGRPGSHGYVRSPSPVSTQETMLQORPSVFQ-----GTNGTSTVITPLDPTA--- 1791  
Db 2492 AAKSEQSPRRHG-PTTPKSALRRD-----FDEGYCGGTTSTLYRTPSKVIRF 2541  
QY 1792 -----QLRIMPLPAGGPSISQGLPASR-----YNTAADALAAALVD 1826  
Db 2542 RDEDEDEQFEAPRPSKQTSQSDRMVGTTHDVLDCLTENTKILQRLKTLADQPSATRS 2601  
QY 1827 AAASAPQMD-----VSKTESKHEAARLEENLRGRSAV-----SEQOOLEQK 1869  
Db 2602 YASSTEGDLGSLRLMRETSPITRTGTHTGVPRLTGTGENIDRLSIKSIKSIDTLCEEK 2661  
QY 1870 TLEVEK--RSVOCLTTSFAFSGKPOPHSSVYVSEAGDKGPPPKPSRYEEELRTRKTTI 1927  
Db 2662 PYQKEKORYIDSLTDSLHPASKSLIEDLSLRS-----LSRSE-----SGRSIH 2709  
QY 1928 TAANFIDVI-ITRQTASDKDAREGSSSSSSLSHRYETPDSADIEVISPASSPAPPO 1986  
Db 2710 RSGDYAPSVIRVTEHRSLSGADSRRSPLCNDRDTSPLHHR--SHRDISRELSP----- 2759  
QY 1987 EKLQIYQEVVYKANAENDPTQYEGPLHHYRPOQESPSQOQLPPSSQAEGMGVQPTH 2046  
Db 2760 -RRRLLEEDEERKDRSSRVRR-DNLLPNYFADNRS-----ELSSGSLTGF-----NH 2807  
QY 2047 RLITLADHICQIITQDFARNVSSOTPOQP-----TSTFQNSPSALVSTPVRTKTS- 2098  
Db 2808 KVDROLETCAKYADD---RSACETPLSHPYESTTATRHSHDTPVOIPNPGASATAT 2864  
QY 2099 ---NRYSPESQAQSVHQRPG-----SRVSPENLVDKSR 2129  
Db 2865 DSFPVPVSPYROPYDHYHRSFGAGGTPLYQPKLEIRHTVTSTFYDRFLTEKQIERQT 2924  
QY 2130 GSRPKSPERSHVSSEPYEPISPOQVYVHEKQDSLILLSORGAEPAQRNDARSPGIS 2189  
Db 2925 HSRP---PSRS-----PVVSPSVPA-----KSYVELCSTSGT-----SSTATSTSTSS 2965  
QY 2190 YLPSPFTKLENTSPMVKSKQEIFERKLNSGGGSDMAAAGPTEIFNLPAVTTSGSVSS 2249  
Db 2966 FMSSSY-----AGFSFSLPSASNFYLNPGSGSGGSISSISPRASCSDLRS-TTSGPTST 3019  
QY 2250 RGHSFADPASNLGLEDIIRKALMGSDFDKVDGHGVMSQPMGVPGTANTSVTSGETRR 2309  
Db 3020 -----STTSVTSYVPYNYFTSFTSRDLNDPII-----TSTTSVAVTSSSLTH 3061  
QY 2310 EGDPSPHSGGVCKPKLISKNSRKSPI 2339  
Db 3062 STGVYNPMMSFTLREPLASSLGGSSASPL 3091

RESULT 22

Q9JLT1 PRELIMINARY; PRT: 4880 AA.

AC Q9JLT1;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE MULTIDOMAIN PRESYNAPTIC CYTOMATRIX PROTEIN PICCOLO.  
OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_taxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20170257; PubMed=10707984;  
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,  
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;  
RT "Piccolo, a Presynaptic Zinc Finger Protein Structurally Related to  
RT Bassoon";  
RL Neuron 25:203-214(2000).  
DR EMBL; AF138789; AAF07822.2; -.  
DR InterPro; IPR000008; -.  
DR InterPro; IPR001478; -.  
DR InterPro; IPR002965; -.  
DR Pfam; PF00168; C2; 1.  
DR PRINTS; PR01217; PRICEXTENS.  
DR PROSITE; PS00499; C2\_DOMAIN\_1; UNKNOWN\_1.  
DR PROSITE; PS50004; C2\_DOMAIN\_2; 1.  
DR SMART; SM00239; C2; 1.  
KW Matrix protein.  
SQ SEQUENCE 4880 AA; 530148 MW; 73951EE4ED83EA68 CRC64;

Query Match 3.7%; Score 462; DB 11; Length 4880;

Best Local Similarity 18.2%; Pred. No. 5,2e-16;

Matches 569; Conservative 398; Mismatches 1066; Indels 1088; Gaps 140;

QY 35 QOEFVAVPYRSHLHVSVASQLLQO--OQQOOLRRRSLSEFHGSDR-----PQERR 86  
Db 1117 EAEKVPPEKETASIEKTPPMVTTDQKLESECKSKVSALEPKPESEKATSAADKKERK 1176  
QY 87 TSYEPFHP-----GPSVD-----HDSLE-----SKRPR 110  
Db 1177 PPAEKPPELEKPPVDPKLLPPEAKPLSSEGEKEHLKAHVQIPEEPTGKVAAKAGE 1236  
QY 111 LEQVSDSHFORVSA--VLP-----LVHPLPEG-----LRASADAKDPAGF 150  
Db 1237 EEQPDSPREALPGATPLTLPKAGEKERAVALPQAEGSSKQGGQGERSKEKTEKEDSDT 1296  
QY 151 GKHEAPSP-----ISGQPCGDDQNASPK-----LSKEELIQ-SMDRVDR 190  
Db 1297 SSSQQPKSPQGLSDFTGSSDGISS-LGEIPLSPSEKDLKGLKDKDSFQESSPSSPS 1355  
QY 191 ETAKVEOILKLKKQOQ--LEEEAAKPEPEKVPSP--PVVEOKHRSIVQ---IYDE 242  
Db 1356 DLAKLESTVLSILEAQASTLYGEAKKKTQPK-ISPKEPQDQKTQTASETLDITISEE 1414  
QY 243 NRKAAE-----AHKIFELGPKVELP-LYNQPSDTKYVHENIKTNQVMRKKLIILFF 293  
Db 1415 EIKESQEKVSPKPDSEQGFPSRKEKPELVDDLSPRASYSVDSEDSSESENSESPVVR 1474  
QY 294 KRR---NHAKRQROKICQYDQIMEAKVKVDRIENPRKAKESKTREYEQFPEI 349  
Db 1475 KRRTSIGSSSDYKQDSQSGSEEDFIRQI--IEMSADEADSGSEDEFIRSOLKEI 1532  
QY 350 RKQEQQRORFORVGORGAGLSATARSHEISEIIDGSEQEN--NEKOMRQLSVIPPM 407  
Db 1533 SGVGSQKREAKG-KGKGVAGKHRLTRKSTSTFDDAGRRHSHHDEDETFDESPELK 1591  
QY 408 F-----DAEQ-----RRVKFINNGLMEDPMKYKDRQFMNVWTDHEKEIFKDKF 452  
Db 1592 FRETQSQSEELVWAGGGLRRFRFTIELNSTADKYSESSQKKTILYFDEPEPEMESL 1651  
QY 453 IOHPK-----NFGLIASYLERKSVPCVLYIYLTKNE---NYKALVRNRYGRGRN 502  
Db 1652 TDSPEDRSRGEGSSSLHASTPTGTSPTSV--SSLEDSDSPSHKKGESQQRKARHS 1709  
QY 503 QOIARPSQOEKVEKEDEKAEKTEKKEKDEEKEKDEKSKENTKEDKIDGTAETE 562  
Db 1710 HGPLPTIEDSSEELREBELLEKQEQKQRELEQOQKSSSKSKKDKDEL--RAQRR 1767  
QY 563 EREQATPRGRKTSQGRKGRITRSMTNEAAAAAATAEPPPLPPPE--PIST 620



```
Qy 2225 DMAAAQPGTELFNPAVTTSGVSRGSHFADPASNLG-LEDIIRKALMGSDDKVEDHG 2283
Db 3678 ----VSTGTQ-----STYSTGTATRRRCRTNTWARAKILQIDREL-----DIVERES 3723
Qy 2284 VMSOPMGVPG-----TANTSVVTSGTTRREEDPSPHSGGVCKPKLISKNSRKSKSP 2338
Db 3724 AKLRKQAELEDEEKEIDAKLRYLEMGINRRKEA-----LLKEREKERAY 3769
Qy 2339 IPG-----QGYLGTERTSPSSVSHSEGDYHRQTPGAWEDRPSSTGSTQF----- 2383
Db 3770 LOGVAEDRDYMSDEVSTRPSRVESQHG-----VERPTAPQTEFSQFTIPP 3816
Qy 2384 -----PYNPLTMRMLSTSTPTPTACAPSAVNOAAPHQONRIREREPAPLLSAQYE 2433
Db 3817 QTQTEANQLVPTSPYTOQYSS-PALPTQ-APTPTQOSHQQOQTLHYQOYSP-----YQ 3869
Qy 2434 T 2434
Db 3870 T 3870

RESULT 23
Q9JKS6 PRELIMINARY; PRT; 5085 AA.
AC Q9JKS6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MULTIDOMAIN PRESYNAPTIC CYTOMATRIX PROTEIN PICCOLO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RT "Multidomain presynaptic cytomatrix protein Piccolo, long splice
variant.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF227534; AAF63196.1;
DR InterPro; IPR000008;
DR InterPro; IPR001478;
DR InterPro; IPR002965;
DR Pfam; PF00168; C2; 2.
DR PRINTS; PR01217; PRICHEXTENS.
DR PROSITE; PS00499; C2_DOMAIN_1; UNKNOWN_1.
DR PROSITE; PS00004; C2_DOMAIN_2; 2.
DR SMART; SM00239; C2; 1.
DR Matrix protein.
SQ SEQUENCE 5085 AA; 552711 MW; 5A1BB543201A7450 CRC64;

Query Match 3.78; Score 462; DB 11; Length 5085;
Best Local Similarity 18.2%; Pred. No. 5.5e-16;
Matches 569; Conservative 398; Mismatches 1066; Indels 1088; Gaps 140;

Qy 35 QOEFAVPYRSHLEVSQASLLQ--QOQOOLRRRPSLLSEFHPGSDR-----POERR 86
Db 1117 EAEKPVPEKETASTIEKTPPWTTTQKLEESGKKSKVSALPEKPSSEKKAISADKKERK 1176
Qy 87 TSYEPFHP-----GPSVD-----HDSLE-----SKRPR 110
Db 1177 PPAEKPPELEKPIPVDKLPPAKPLSSEGEKEHILKAHVQIPEEPTGKVAARAGE 1236
Qy 111 LEQVSDSHFQVRSAA---VLP-----LVHPLPEG-----LRASADAKDPAG 150
Db 1237 EEQPDPSRPEALPGATPLTLTKAGEKRAVAQQAEGSKDGQERSKEKTEKEEDKSDT 1296
Qy 151 GKHEAPSSP-----ISGPGCDQONASPSK-----LSKEELIQ-SMDRVD 190
Db 1297 SSSQPKSPQGLSTGTSGSGISGS-LGEITSLIPSEKDLLKGLKDKDSFQSSSPSS 1355
```

```
Qy 191 BIAVEQOILKKKQOO--LEEAAKPEPEKPVSP--PPVEOKHRHSIVO-----IIVDE 242
Db 1356 DLAKLESIVLSLEAQASTLVGEKAEKKTQPK-ISPKEPQDOOKTQTASETLDTITISEE 1414
Qy 243 NRKKAEE-----AHKIFGLGPKVPEL-LYNQPSDTKVYHENIKTNQVNRKKLLILFF 293
Db 1415 EIKSQEKKVSPKKDSEQFPSPRKEHKEKPELVDDLSPRRASYDSVESSESSESPVVR 1474
Qy 294 KRR---NHARKOREOKICORYDQOLMEAWKVKVDRIENPRKAKESKTREYEQFPEI 349
Db 1475 KRRISGSSSDEYKQDSQGSSEEDFIRKQI--IEMSADEDSGSEDEFFIRSQLKEI 1532
Qy 350 RKQEQQRFRVQORGAGLSATARSSEHSEIIEIDGSEQEN--NEKOMROLSVIPPM 407
Db 1533 SGVGESOKREEAG-KGKGVAGKHRLTRKSTSFDDDDAGRRHSHHDEDETFDESPELK 1591
Qy 408 F-----DAEQ-----RRVFINNGLMEDPMKYKORQFMNVTWDEHEKTFKDKF 452
Db 1592 FRETQSQSEELVAVAGGGLRRFKTIELNSTADKYSSSQKKTILYFDEBEPELEMESL 1651
Qy 453 IQHPK-----NFGLIASYLERKSVPCVLYLYLTKNKNE---NYKALVRNYGKRG 502
Db 1652 TDSPEDRSRGEGSSSLHASTPTGTSPTSV--SSLEDSDSPSHKKGESQOQRKARHS 1709
Qy 503 QOIARPSOEERKEEDKAEKTEKKEEKKDEEKDEKSKENTKEKDKIDGTAEETE 562
Db 1710 HGPLLPTIEDSSEBELREBELLEKQEKQKQKQKQKSSKSKKDKDEL--RAQRR 1767
Qy 563 BREQATPRGRKTANSQGRKGRITRSMTNAAAASAAAAATEEPPPLPPPE--PST 620
Db 1768 ERPKTTP-----SNLSPIEDASPEEL-RQAAMEBELHRSSCEYSPSIESDPGEFISP 1821
Qy 621 E-----PVETSRW--TEEMEVAKK----- 639
Db 1822 EKIEVQVYKLPAAVSLYPTDQSVMQKGVKALSAEEMEMQKPKHYKAPAA 1881
Qy 640 -----LVHGRNMAIAKMWGTKSEAQCKNFYFNKRRHN 674
Db 1882 NERDEVFEKPLYGMLIEDYIESLVEDTYNGSVGDSLL-TRQEEQ--NGFMQOGRQ 1938
Qy 675 LDNLQO---HKQTSRKPREERDVSOCEV-----ASVVS 707
Db 1939 KVRLOEQIYDDPMQKISDLQKEFVELESLSVVPQEDIVSSYIIPESHEIVDLGSMVMS 1998
Qy 708 AQEDE---DIEASNEE-----ENPEDEVEAVKP 733
Db 1999 TSEKKLLDADSAYEELMRQOVQVTDGSSPVQTTIGDDMAESTLDFRVQDASLT 2058
Qy 734 SEDSPENATSRNGTE-PAVEL-----E 754
Db 2059 SGASLTDTSSATLSIPDKITQOFSAELEDEYVTDYTRIQDIIAHESLILTYSEPSE 2118
Qy 755 PTTETAPSTSPSLAVP-STKPAEDESQVNDNISASATAQOMVDQOEHAEE----- 807
Db 2119 SATSVPPSDTEPLTSSISVCTDSSSPVTTLSLTTVTTEPADVMTKFKDSEISSTFY 2178
Qy 808 -GSVCD-----PPKATKADSVDE 825
Db 2179 PGSIIDPIDISVLDRTIMPESRTNEDRIVLSFGMAPSVVESVGTFRPQADTISTD 2238
Qy 826 VRVPEN---HASKVEGDNTERDLDRASEKVEPRDEL-----VVAQOI 866
Db 2239 LPISEKDLIKKKTGDGIIILEVLDAVKDKRESEAEELTKISLPEPLQAQAPSSVTAPOI 2298
Qy 867 NAQRPESQNDSSATCSADEVDGEPERQRMFMDKPSLLNPTGSLVSS----- 918
Db 2299 KEQHVSPHS-----VSGKISGQ-----EKTPLPSGSLPVSTPHSPSKSRPF 2339
Qy 919 -----PLKNPDLPLPOLQHRAAV--IPPMVSCTCPNIP-----IGTPVSVGA 958
Db 2340 FRSSSLDISAQPPPPPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 2399
Qy 959 LYQRHIKAMHESALLEQORQEQIDLECRSSTSPCGTSKSPNREWEVLQAPHOLITNL 1018
```

Db 2400 IVTTHVDAL---TWVEAAARRS-----NGLPATKCAIAPP-----VPPKPSQIPTGL 2446  
Qy 1019 -----PEGVRLPTTRTPPPPLIPSSKTVAASEKSPFIMGSSISQGTGGVYLTSHNQAS 1073  
Db 2447 VFTHRPEAIKPIA-----PXPAPV-QIPVTTQRP----- 2475  
Qy 1074 YTOETPKPS-----VGSISGLPQROE-SAKSATLPYIKOEFSRQNSOPE----- 1120  
Db 2476 -TDTCPKPTGLSLTMSLSNLVTSADYNVPSPTSP-----LSPHSNKSPPRYKSLMDT 2528  
Qy 1121 -GLLVRAQHEGVVWGRTAGAIQEGSTTRTPSK-LSVESIPSLRGSITQGTPALPQGTGP 1178  
Db 2529 YVWITLPEPQGTPTDSSAAQAITSMPLGSPKDLVSLTFVSF-----VPPMTSTEIP 2581  
Qy 1179 TEA-----LVKGSISR-MPIEDSSPEKCREAASKGHVYI 1212  
Db 2582 SASQPTLYTSGALGTFSTPAVTSALFTVPTLSLQFLPAEASRPEVSAVSAPV----- 2636  
Qy 1213 EGKSGHILSYDNIKNAREGTRSPRTAHEISLKRSYVESVEGNIKQGMMSRESVPVSAPLEG- 1271  
Db 2637 -----SVAPRSVSIPIPEPLALDRHQYKENGKL-----PLIGD 2670  
Qy 1272 -LICRALPRGSPHSDLK--ERTV-LSGSIMOGTPRATTESFEDGLKYPKQIKRESPPPIRA 1327  
Db 2671 AIDLRTIPK-----SEVKTERKCMDSASAMDVKQTTANEV-----YRQISAVQPSIIN 2721  
Qy 1328 FEGAITTKPKPYDGIITIKEMGRSIEIPRODILTOESRKTPEVWQSTRPIIEGSSISQGT 1387  
Db 2722 LSASSLGT-----VTMDS-KTAVAVTCT-----DTT 2748  
Qy 1388 IKFDNNSQSAIKHNKVLITGPKSLRGMPPLEIVPENIKVVERGYEDVKGAVETVRSR 1447  
Db 2749 IYTTGTEQVGEHAVTS-----PLQL-----TTSK 2774  
Qy 1448 HTSVVSSGSPVLRLSTLHEAPKAQISPGIYDDTSARTPTVSTQNTMSRGSPPMNRSTDVTI 1507  
Db 2775 HTELP-----YRKPSQAFPTIRDE-----APINLSLGFSAQAQAVTLAVTKPVT 2818  
Qy 1508 PNKSTNHERKSTLTP-----TQRESIPAKSPVPGVDVVSF-----SPDPH 1550  
Db 2819 PPVGVTNGWTSTLSQGVADGEVDLSISKSHRTVVTHDETSNVVTKLIEDEKPYD- 2876  
Qy 1551 HRGSTAGE-----VYWSHLPTQLDPAHPFHR-----ALDPAAAA-----YLFQR 1589  
Db 2877 ---LTAGRAVCCDMVY-----TLPEGRSCTAQOAPATTLPEDRFGYRDDHYQDR 2923  
Qy 1590 QLSPTPGY-----PSQ-----YQLYAMENTROTILNDYITSQOMQVNL---RPD 1630  
Db 2924 --SGPYGRGIGGKMKPSMDSNLPAGHFFYKSKNA-----FDYSGGTGAADVLTSGRVS 2976  
Qy 1631 VARGLSPREQPLGLPYPATRGITIDLTNNPPTILVPHPGGTSTPPMDRITYTPGTQITPPP 1690  
Db 2977 TGEVMDYSKTTG-PYPETROVISG-----GISTPQYST-----ARLPPP 3017  
Qy 1691 RP-YNSASMSGPHTHLAAAASAEERERERERERERERERERERERERERERERERERER 1746  
Db 3018 GPQXGVGSV-----LSSNGVYSSVATPIPSRFA 3047  
Qy 1747 ---RGS--HGVRS-----PSPSVRTQETMLQO---RPSVFOQTNGTS-----VITPLDPT 1790  
Db 3048 ITTQGSIFSTVTRDLSGIPITDAMTSLALHQSOPIPRSFYITGASETIVTGIDIN 3107  
Qy 1791 AOLRIIMPLPAGGPGSISQGLPASYRN-----TAADALAALV-----DAAAASAPQM 1834  
Db 3108 ASLQITIM-----ETLPAETWDSVPTLTLTASEVSEVGEESTLLIVDEDKQOOL 3159  
Qy 1835 DVSK-----TESK 1843  
Db 3160 DLERELLELEKIKOORFAEELEWERQEQIRFEQEKIMVQKKLELOSQMOKHLLYQOEE 3219  
Qy 1844 HEAARL--EENLRSRAVSEQOLEQKT-LEVEKRSVQCLYLTSSAPSGKPPQHS----- 1896

Db 3220 ROAFMMRQETLAQQQLQLEQIQLOOQLOOQLEBQKLRQIYQYNDPSGTSSTPQTTEQ 3279  
Qy 1897 SWVYSEAGKDKGPP-----PKSRYEEELRTRGKT-----TI 1927  
Db 3280 AILLEGYAAATSGSFQWATEDATTASTVVAIEIPQSQGWYTVQSDGVTOYTAPPGLISV 3339  
Qy 1928 TAANFIDVITRQ-----IASD-----KDARERGSQSDSSS-- 1959  
Db 3340 SEIPLTDVVVEEKQPKKRSSGAKVGOYDEMGESVADDPRLNKKIVDSGVQTDDEETAD 3399  
Qy 1960 -LSLHRYETSDALEVSPASS-----PAPPOEKLOTQYQEVVVKANQANENDPTROYE 2011  
Db 3400 RSYASRRRTTKKSDVTSYQTDDEQDEMDPFSRKRKARTGK---YGDSTAEQDKT--- 3453  
Qy 2012 GPLHHRPQOQSPSPQOOLPSSOAGMGQVPRHRLITLADHICQI1-----TODFAR 2065  
Db 3454 -PL-----SKVSSVAQIVAEISVQTEPVGTI-RTPSIARVDKAVELLIKHISAPEKTYG 3507  
Qy 2066 NOVSSQTPQOQPTSTFQNSGALYSTPVTRKTSNRYSPESQAQSVHHQR-----PGS 2117  
Db 3508 GSLGQOTETSDT---QSPPYLGATSPKDK--KRPTPLEIGYSSSHLRADPTVQLAPSP 3562  
Qy 2118 RVSPENLVDKRSGSRPKSPERSHVSSEPYE-----PLSP-----POVPVVEHKQDS 2164  
Db 3563 PKSPKVLVSPISPLSPGNALPAFV---PYEKLPLDDISPOKVLHPDMAKVPASPRTAK 3619  
Qy 2165 LLLLSORGAEPAEQNDARSFGSISYLPSPFTKLENTSPWVSKKOEIFRKLINSSGGGDS 2224  
Db 3620 MMQSRMSDPKPLSPADESSRAPFOYSEGFTTKGSQT--MTASGTQKKVKRTPLPNPPPEE 3677  
Qy 2225 DMAAAQOQTEIFNLPAVTVSGSVSRGHSFADPASNLG-LEDIIRKALMGSDDKVEDHG 2283  
Db 3678 ---VSTGQ---STYSTMGTFASRRRCRTNTMARAKILQIDREL-----DLVERES 3723  
Qy 2284 VMSQPMGVPG-----TANTSVTSGTTRREBGPSPHSGVCKPKLISNSRKSXP 2338  
Db 3724 AKLRKKQAELEDEEKEIDAKLRYLEMGINRKEA-----LLKREKREKRAY 3769  
Qy 2339 IPG-----QGVLTGTERPSSVSVHSEGDYHROTPGWAWEDRPSSTGSTQF----- 2383  
Db 3770 LQVADRDYMSDESEVSTRPSRVESQHG-----VERPRTAPOTEFQFIPP 3816  
Qy 2384 -----DYNPLTMRMLSTPPTPIACAPSAVNQAAPHQONRWRPREPAPLISAQYE 2433  
Db 3817 QTQTEAQLVPPTSPYTOYQYSS-PALPTQ-APTPTVQOSHFQOQTLYHQVSP-----YQ 3869  
Qy 2434 T 2434  
Db 3870 T 3870  
RESULT 24  
Q9VC00 PRELIMINARY; PRT; 2768 AA.  
ID Q9VC00  
AC Q9VC00;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE CG13648 PROTEIN.  
GN CG13648.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,



```
QY 1344 IKEMGRSHEIPRODILTO-----ESRKTPEVVOFPRPIEGSISQGTPIKFDNN---SG 1395
Db 1465 TE-----ADISTEPSAWEKEASGETSESDNEIDAGASSTPPVPSADEDKTST 1513
QY 1396 QSAIKHNKSLITGPKSLRGMPPLEIVPENIKVVERGYEDVKAGETVRSRHTSVSSG 1455
Db 1514 EKTVEADDDKFTTAVPLAGDEESNLPKLPQDI-----FEB-----EAPVAVTTAA 1558
QY 1456 PS-----VLRSTLHEAPKALSPGIIYDTSARTPVSYONTM-----1492
Db 1559 PSKDGEQKPVVEKEPIEDGQKPIEDETS---TPTSSENEIEPESDRATTIAPKEEPS 1615
QY 1493 --SRGSPMMNRSTSDVTPNKNSTHERKSTLPTQR-----ESIPAKSPVPGVDVP 1541
Db 1616 EFTSGAP---TKDEPAEPSTADPESDSKETPESEVPTTVAPAGEKIPSTISITDEEPT 1671
QY 1542 VSHSPF---DPHHRGSTAGEVYWSHLPTQLDPAMPFHRALDPAALYLFQRLSPTGYP 1598
Db 1672 ATSAPVAKPDEDEKETSTEI-----PTDA-----PASSE---EDENSSTDQIP 1712
QY 1599 SOYQLYAMENTQITLINDYI-----TSQOMVNLRPDVARGLSPREOPLGLPYPATR 1650
Db 1713 SEVPEKKPTPAQTPEEGDIVGATAAPTSDV-----PPVQR--LPEEVLAEIPQSTE 1765
QY 1651 GIIDLTMNPPTILVPHPGGTSTPPMDR---ITYIPGTQITPPP-----RPYNSASNSP 1700
Db 1766 -----TGIKQOEDTAAPSIDREPEVTEDEATTVAPISEKDEKTEEEKPVE 1815
QY 1701 GHPTLAAASAERERERERERERE-RIAAASDLYLRPGSEQPGRPGSHGYV-----1754
Db 1816 QKPT--GEEPSBEEKEKPIEQDVSTEGVPVSTEASEAGSTESSEV-KPSTGEVAEKPE 1872
QY 1755 -RSPS-----PSVRT-----QETML---QORPSVF-QGTNG 1780
Db 1873 DKQPSSTAQAVETIPELSTELPAODGDKPSEAPVDSDEDTAPSDEKIPSVSGEEVEG 1932
QY 1781 TSVITPLDPTAQLRIMPLFA-GGPSISQGLPASRY-----NTAAD-----ALAAALVDA--1828
Db 1933 PVTVTASQAEEDELKTPAEPSEPSSTDKVPETEYQKPEDETKADETPESVTQVSDVATS 1992
QY 1829 ASAPOM--DVSKTKESKHEAARLENLRSNAVSEQQOQLQK-----TLEVEKRSVQCILY 1882
Db 1993 TSAPVAGVDIEKDEQATASPEEEIEKPTIAPAAEIPQSEKEPVDEQEVESGKTATPA 2052
QY 1883 TSSAPPKGPPOPHSSWYSEAGKDKGPPKSRYEELRGTGTTITAA-----NFID 1934
Db 2053 ESDGQPIDEIAPATSGPIDEA--STAATK-----EE-----STTVASAASPAVHDEIKD 2101
QY 1935 VIITRQIASOK-----DAREGSSQSDSSSLSSSHRYETPS 1970
Db 2102 VTTTPVADAEKEVAAPQDETSTSIDVSTDSPTAQDDEKQDKTEAPVAPTIVSSPTADSA 2161
QY 1971 DA----IEVISPASSPAPQEKL-QTQPEVVKANQANDEPTROYEGLPHHYRPOQE--2022
Db 2162 DSSTPEVPSPEVIDTKPMDIMSQTTAPHT--ADGAASTSTEDQOAPVTVSPQDAEK 2219
QY 2023 ---SPSPQ--QOLPPSSQAEQGVPRP-----ADGAASTSTEDQOAPVTVSPQDAEK 2047
Db 2220 TPVSPAQODSKTPSSPAQDAEIPATATPLDNDKIPATVAPQTDGCVPAATAPLDEDK 2279
QY 2048 LITLA-----DHICQIITQDFARNQVSSQTP-----QOPP-----2077
Db 2280 IQTTAAPLDEEKIPSTAAPLDDKIPAPVPVFDVPSSEKPAVSEYDGEESTEPVPHD 2339
QY 2078 --TSFQNSPALVSTPVYTKTSNRYSPESAQSVHGORPSRVPENLVKSGRSRPGK 2135
Db 2340 VETSTDEPSTDAKLKPPTSAPATPSESATEAEIV-----PETAAPLEKEVPEK 2389
QY 2136 SPERSHVSESE-PYEPISPPQVPVVEHEKQDLSLLLSQGAEPAEQNRDARSPTSILPSF 2194
Db 2390 ATEQPELEKETPERKATEQPEL-----EKE-----TPEKATEQP--2422
```

```
QY 2195 FTKLENTSPMKVKOEIPLFRKLNSSGGSDMAAQAOPGTEIFENLPAVTTSGSVSRGHSE 2254
Db 2423 --ELEKETPEKATEQPELEKEV-----TDKATEQPE-----SVDEKTP-----2459
QY 2255 ADPASNLGLEDIIRKALMGSDFKDVEDHGVMSQPMGVVPGTANTSVVTSGTTRREEGDP 2314
Db 2460 -----EPVVKPSL-----DSTEDEESVESEESADKKDKNE--TEEDTDKKHEEP 2504
QY 2315 SPHSGGVCKPKLIKSNRKSRSKSPIGOCY-----LGTERPSSVSSVSHSGDYHROT 2367
Db 2505 E-----VPVAVSEI-PQPSSEAVPTTGHPLFHLASSTTTTPPAVDORVGE-----2548
QY 2368 GWAWEDRPSSTGSTOPYNPPLTMRMLSSPTPTPIACAPSAYVQAAPHQONRIWEREPAPL 2427
Db 2549 ---EDEENTTVK-----LSSSTTTTSPVTSAPSTTTTVAOQOQ-----PITPPY 2593
QY 2428 LSA-QYETLSDSD 2440
Db 2594 GHAPYEDEYDEE 2607
RESULT 25
Q9UHA8
ID Q9UHA8 PRELIMINARY; PRT; 2296 AA.
AC Q9UHA8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE SPLICING COACTIVATOR SUBUNIT SRM300.
GN SRM300.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RA SEQUENCE FROM N.A.
RA Blencowe B.J., Bauron G., Eldridge A.G., Issner R., Nickerson J.A.,
RA Rosonina E., Sharp P.A.;
RT "The SRM160/300 splicing coactivator subunits.";
RL RNA 0-0(2000).
DR EMBL; AF201422; AAF21439.1; -.
DR InterPro; IPR002965; -.
DR PRINTS; PR01217; PRICHEXTENSN.
SQ SEQUENCE 2296 AA; 251962 MW; 17C0BD4EA10A9CF9 CRC64;
```

Query Match 3.6%; Score 453; DB 4; Length 2296;  
Best Local Similarity 19.6%; Pred. No. 5.7e-16;  
Matches 504; Conservative 288; Mismatches 979; Indels 796; Gaps 114;

```
QY 44 RSSHLEVSAQSLLQO-QQQQLRRRRPSSLLSEFHPGSDRPOERRTSYEPFHPGSPVDHD 102
Db 305 REMLSVNOVLPTQIRASSPETATKQPS-----SPYEDKDKKKEK-SATRPSFP-----354
QY 103 SLESKRPRLEQVSDSHFORVSAAVLPLVHPLPEGLRASADAKDPAFGGKHEAPSSISG 162
Db 355 -----ERSSTGPEP---PAPTLLAERHG-----GSPQPLATPLLSQ 388
QY 163 OPCGDDQNASPSK-----LSKEELIQSDMDRDREIAKVEQOILKLLKKQOOLEEAAKPP 217
Db 389 EPVNPPEASPTDRSPKSPKPLQSS---SSESSPSPQTKVSR-----HASSP 438
QY 218 EPEKVPSPPPVQKHRSITVQIYDENRK-----KAEAAKIFEGGLPKVELPLYNQPSDK 273
Db 439 ESPAPA---PAPGSHREISSPTSKNRSHGRAKRDKSH-----SHTPSRMGRSRSPATAK 491
QY 274 VYHENIKTNQVMRKLILFFKRRNHARKQREKICQRYDQLMEAWKVKVDRIENNRKA 333
Db 492 RGRSRSTP-----TKGRHSRSRSPQ---WRRRSQAQRWGR-----SRSPQRRG 532
QY 334 KESKTREYEEKQFPIRKQROQOERFORVGORGACLSATIRSEHEISEIIDGLSEQENN 393
Db 533 RSR-----PQPGWSRSRNTQRRGRSRARRGRS-HSRSPATRGRSRSTP 578
```











```
Db 4248 VSESEPKPEPSVEQTVKKRKPSTFADEPATEIVIKESKPAEVVTDHAHIKTKPKKK 4307
QY 1471 LSPGIYDTSARRTPVQNTMSRSPMMRTS-DVTTPPNKST-----NHERKSTLPT 1524
Db 4308 VT-----DVEAEELIKITEEYQPOBILIEEVESEVITETKKTAPVVEERTYKIGIKET 4362
QY 1525 QRESIPAKSPVGVDPVYSHSPDFPHRGSTAGEVYWSHLPTQLDPAMPFHRALDPAAAA 1584
Db 4363 EPEK-PAEAIVEEPEPVTE-----PIEA-----4386
QY 1585 YLFORQLSPTGPGYQOLYAMENRTQILNDYITSQMQVNL-----RPDVAGLSPREQP 1641
Db 4387 -----PKPEVEEHKVRVTEETPRELVEEVEEVEEVRKVRIRKKPKPEI-----KEEP 4432
QY 1642 -----LGLPYP-----AVRGIIDLTMNP-----TILVPHGCGTSTPPMDRITYIP 1682
Db 4433 EAEVTVSTPKPVEEVEATSIAPVPEQTEEEAADLKITII-----EETPPQELVQIE 4487
QY 1683 GTQITFPRPYNASMSGPHPTHLAAAASAERERERERERERERERERERERERERERER 1742
Db 4488 EIEIVEEKAPEE-----QPTDFTFKDKSEKKTVEELPEQVTIQKKKK-----APVP 4538
QY 1743 EOPGRGSHGYVRSPSVR-TOETMLOORPSVFGTNGTSVITPLDPTAQLRIMPLPAG 1801
Db 4539 EVVEEPEAEFLVKPTPVQVTEEAKI-----TKSKKPVKEEAAALKV-----4583
QY 1802 GPSISQGLPASRYNTAADALAALVDAASAPQMDYSKTESKHEARLEENLRSAAVS 1861
Db 4584 --TITEEIPTE--PEVQELIEEIEEIEEIEEIEEIEEIEEIEEIEEIEEIEEIEEIEE 4624
QY 1862 EQOOLEQKTELEKRSVQC--LYTSSAPPSGPKPQPHSSVYSEAGKDKGPPPKRYEBEL 1919
Db 4625 ---VEDKEVSLPKKKKRAPIVEEPEAEITLKPVKSEEVQEEAKIVKKPKKI---DEV 4677
QY 1920 RTRGKTTTANFI--DVIITRQIASDKARERQSSDS-----SSLSHRYETPSDAIE 1974
Db 4678 AVADELTVKVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 4737
QY 1975 VISPASSAPPOEKLOTYQ-----PEVVKANQANDPTROYEGPLHHRYPQOESPS 2025
Db 4738 VV--AEVILPKKAKTEVTEEBEFSVDVLPKPKKVTTEESDQTVQL-----KKKK 4785
QY 2026 POQLPPSSQAEQMGQVPRHLITLADHICQITQDFARNQVSSQTPOQPTSTFQNSP 2085
Db 4786 PKPVEEAADELKQTVVEERPVIEEE--EIEEAV---VIRKKPKKPEPTVEDLE 4839
QY 2086 SALVSTPVRTKTSRYSPESQASQVHHRPGSRVSPENLVKSRGSRPKGSPERSHVSE 2145
Db 4840 ETEFSLSPKKPHTINEGVE-EAATVLLKRP---VRPTTL-DEAAAEELSIKROEEYEGE 4894
QY 2146 PYEPI-----SPPOVPVVVHEKDS 2164
Db 4895 DIEEFVVSQQRKPKPLQITEEDEEA 4919

RESULT 28
QSN541
ID QN541 PRELIMINARY; PRT; 4900 AA.
AC QN541;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE Y2C2A.1 PROTEIN.
GN Y2C2A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
```

```
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RC SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006731; AAF60483.1; --
SQ SEQUENCE 4900 AA; 551244 MW; D837F46317C902D5 CRC64;

Query Match 3.6%; Score 449.5; DB 5; Length 4900;
Best Local Similarity 19.6%; Pred No. 2.5e-15;
Matches 584; Conservative 388; Mismatches 1098; Indels 917; Gaps 139;

QY 24 SVQYTFPNTRHOOEFAVPDYSRHLE-----VSQASQLLOOQOQQLRRRPS 70
Db 181 STENEITNTSQTANIVDPGSDIDSEENKLAALKEESILKASQEAASDNHEKERSAT 240
QY 71 LLSEPHGSDRPOERTSYEPHGPSPVDHDSLESKRRLQVSDSHFRQV-----122
Db 241 SKADY-----ERSFDQDVTYERSPLLEPSEEPVMSKEPELTQEEIDHIARIQSLAEQSS 296
QY 123 --SAAVLPLVHLPEGL-----RASADAKDPAGGKHEAP---SSPTS 161
Db 297 FQOASTIPDRPPLPARLPTVDEPIVISEQHEEDRSATSGADYERSFDQEVYERSPLL 356
QY 162 GPCGDQDNASPKLSKE--ELIQSMDRVREIAKVQOILKLKKQOOLEEAAKPEPE 220
Db 357 -EPCE-----PVMSKEPELTQ--EIID-HIARIQ-----LAEKSSFEQASTIPDRPP 402
QY 221 KPVSPPPVEQKHSRIVQIIIDENRKKAEAHKIFEGLGPKVELPLYNQSDTKVYHENIK 280
Db 403 LPARLPTVDEP--IVISEQHEEDRSATSGAD-----YERSFDQDVTY--K 445
QY 281 TNQVVRKKLILFFKRRNHARKQEQICQRYDQLMEAWE--KVVDRIENPRRKAKESTR 339
Db 446 SSPLEPSEDVMSKEPELTQEEIDHIARIQSLAEKSSFEQSTTIPDRPPLPARLPTVD 505
QY 340 EYEEKQFPIRQORQOQORFVG-----ORGAGLSATIARS-----376
Db 506 E-----PIVSEQHEEDRSATSGADYERSFDQDVTYERSPLLEPCEPVPMSKEPEL 559
QY 377 -RHEISEI--IDGLSDEENNEKQMRQLSVIP-----PMFDAEORRYKFIN 419
Db 560 TQEEIDNIAWIOSIAEQSFE---QASTIPDRPPLPVRLPTVDEPIVSEQOEEEDRSSA 615
QY 420 MNGL-----MEDPMKVYKDRQFMNVMTDHEKEIFKQKFIQHPKN 458
Db 616 TSGADYERSFDQDVTYERSPLLEPSEDVMSKEPELTQEEIDHIWI---QSIAEQSS 672
QY 459 FGLIASYLERKSVPCVLYIYLTKNNTYKALVRRYNGRRGNQOIARPSQBEKV---E 515
Db 673 FEQASTVPDRPPLP-----VRLPTVDEPIVSSE 700
QY 516 EKEEDKAEKTEKEEKKDEE-----KDEKEDSKENTKEKDKIDGTAETE 562
Db 701 QHEEDRSATSGADYERSFDQDVTYERSPLLEPCEPVPMSKEPELTQEEIDHIWIOS 760
QY 563 EREQATPRGRKKTANSQGRKGR-----ITRSMTEAAAAAASAAAA---ATPE 606
Db 761 SAEQSSFEQASTVPDRPPLPVRLPTVDEPIVSEQHEEDRSATSGADYERSFDQDATTY 820
QY 607 PPPPLPPPEPISIEPVEFSR---WTEEME-----VAKKGLVEH-----GR 645
Db 821 RSPLEPCE-----EPVMSRVPPELTQEEIDHIARIQSLAEKSSFEQSTTIPDRPPLPAR 876
QY 646 NWAATAKMYCTKSEACKNFY---FNYKRRHNL-----NLLQHKOKTKSKPPEERD 695
Db 877 LPTVDEPIVSEQHEEDRSATSGADYERSDQEVYERSPLLEPCEPVPMSKEPELT 936
```



Q10465  
ID Q10465 PRELIMINARY: PRT: 7962 AA.  
AC Q10465;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE TITIN, SKELETAL MUSCLE ISOFORM (EC 2.7.1.-) (CONNECTIN) (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-SKELETAL MUSCLE;  
RX MEDLINE=96026330; PubMed=7569978;  
RA Labelit S., Kolmer B.;  
RT "Titins: giant proteins in charge of muscle ultrastructure and  
RT elasticity".  
RL Science 270:293-296(1995).  
CC -!- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE  
CC ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF  
CC SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.  
CC -!- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE  
CC PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE  
CC DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY  
CC ONE TISSUE.  
CC -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.  
CC -!- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE  
CC KINASES.  
CC -!- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 90  
CC IMMUNOGLOBULIN C2-LIKE DOMAINS.  
DR EMBL: X90569; CAA62189.1; -;  
DR HSSP: P56276; 1TLK.  
DR InterPro: IPR003006; -;  
DR InterPro: IPR003598; -;  
DR Pfam: PF00047; 1g; 59;  
DR SMART: SM00408; Igc2; 1.  
KW Muscle protein; Cytoskeleton; Structural protein; Phosphorylation;  
KW Serine/threonine-protein kinase; Alternative splicing; Repeat;  
KW Immunoglobulin domain.  
FT NON\_TER 1  
FT DOMAIN 5618 7792 GLU/LYS/PRO/VAL-RICH.  
FT NON\_TER 7962 7962  
SQ SEQUENCE 7962 AA; 883018 MW; B85240533CBAD58 CRC64;

Query Match 3.5%; Score 444; DB 4; Length 7962;  
Best Local Similarity 18.3%; Pred. No. 9.5e-15;  
Matches 463; Conservative 332; Mismatches 934; Indels 804; Gaps 111;

Qy 34 HQQFAVPD---YRSSHL--EVSOASQLQOQOQOQRRRLSLLSEFHGSDRPO-ERRT 87  
Db 5901 HTKMWVISEKMFASHTEEEVSVTVPEQVEITEEKIHVAVSKRVPPKVPPELPEKP 5960  
Qy 88 SYEPFHGPGSDVHDLSKSRLEPRLEQSDSHQFVSAVALPLVHPLPEGLRASADAKDP 147  
Db 5961 APEEAVPPIP---KKVEPPAPKVPPEVK-----PVPEKKPVPVPKPEP 6003  
Qy 148 AFGGK-HEAPSPTISGQ----PCGDDQNASPSKLSK-----EELIOMDRVDR----- 190  
Db 6004 AAPKVPPEKVPPEEKIPVPVPAKKKAPPAKVPQVQGVVTEKITIVTQRESPPPA 6063  
Qy 191 --ETAKVEQIILKLLKKQOQLEEAAPKPE---PEKPV-----SPPP-VE 229  
Db 6064 VPETPK--KKVPEKKVPKREVEEVPKVPALPKPKPEEKVAVPVPVPAKKAPPPRAE 6121  
Qy 230 QKHSVQIYIDENKKAEEAHKIFEGLPKPKPELYNQPSDTK-----VYH 276  
Db 6122 VSKKTIV---EKKRFAEE--KLSFAVQVQVETRVHEVSAEEESYSEEEGVSYIVR 6175  
Qy 277 ENIKTQ-----VMRKKLILFFKRRNHAKRQKTCQRYDQMLEAME 319  
Db 6176 EEREEREAEVTEYEMEEPVEYVVEKHLHISKR-----VEAPPAEVTQEKKIVLKP 6231

Qy 320 KKVDRITENPRKAKESKTRYEYKOFPEIRKQREOQERFORVGQAGLSATISARSEHE 379  
Db 6232 KIPAKIEPPPAKVPPEAKKIVPEKKVP-----ADVPKKEKV 6268  
Qy 380 ISEIIDGLSEQENNEKQMRQLSVIPPMFMDAEQRVKFFINMGIMEDPMKVYKDRQFMVY 439  
Db 6269 PPKVPPEEKPKVPPEKK-----VPEKVIK-----MEPLPAKVTEKHMQI 6308  
Qy 440 WTDHEKEIF-----KDKFIQHPKNGLIASYLKRSVPCVLYLYLTAKNENYKAL 490  
Db 6309 -TQEEKVLVAVTKKEAPPKARVPEEK-----RAVPEEKVLKPKREEPPEPPAK 6356  
Qy 491 VRRNYGKRRGRNQIA--RPSOEK-----VEEKE-----EDKAETKEKEEKKDEE 536  
Db 6357 V-TERKRVVVEKESIEAPKREPOPIKEVTIMEEKERAYTLEEEAVSVQ-REEEVEEVE 6414  
Qy 537 EKDEKE-DSKENTKEKDKIDGTAETEEREQATPRGKTANSQGRKRGRITRSMTHAAA 595  
Db 6415 EYDYKEFEYEYTEYDQY-----EVEEREY-----ERVE 6445  
Qy 596 ASAAAAATEEPPPLPPPEPISPEPVSRTKTEEMEVAKGLVHGHNWAAIAKMGV 655  
Db 6446 EHEEYITEPEKPIPKVPPEEPVPTKP-----KAPPAKVLKKAPEE-KVPVPPIPKL- 6497  
Qy 656 TKSEAQCKNFYENYKRRHNLNLLQHKQKTSRKPREERDVSOCEVASTVSAQDEDEIE 715  
Db 6498 -----KPPPKVPPEPKKVF-EKIHISITKREKQV- 6528  
Qy 716 ASNEENPEDSEAVKPSSEDSNATSRGNTPEA--VELEP-----TTETAP-----STSP 765  
Db 6529 -----TEPAKVPMPKRVVAAEKVPVPRKEVAP 6557  
Qy 766 SLAVPSTKPAEDESVEYQVNDISAEATQMDVDQOEHSAEGSVCDPPPAKADSDV- 824  
Db 6558 PVRVPEV-PKELEPEEVAEEVTHVEEYLYVEEEETHEEEEF-----ITEEYVPI 6611  
Qy 825 EYRVENHASKVEGDN-----TKEROLDRAS-EKVEPRDEDLVAAQIINAQRPQSDMD 878  
Db 6612 PVKVEPVKPKVPPEKKVPVPPKKEAPPAKVPPEKPEKVPVLPKPEKPPPAKVPPE 6671  
Qy 879 SSATCSADB-----DVGEPERQRMFMDSKPSLLNPTGSI----- 914  
Db 6672 VPKKVPPEKVPVPPKVPPEKVPPEKVPPEKVPPEKVPPEKVPPEKVPPEKVPPEK 6731  
Qy 915 LVSSPLKNPLDLPOLQHRAAVIPPVW--SCTPCNIPICTPVSGYALYORHKAMHESAL 972  
Db 6732 LIPEKKKPTP-----VPKKEAPPKVPKKEPVPVPPVAPALPQEEVLFEEI-VPDEEVL 6785  
Qy 973 LEOQROEQIDLECRSSTSPCGTSKSPNREWEVLQAPAPHOLITN---LPEGVRLPTTR 1028  
Db 6786 PEEVEVLPEEEVLPPEEEVLPPEEEIPEEEV--PPEEVPVEEEVPEEEVLPPEVK 6843  
Qy 1029 PTRPPPLIPSKTTVASEK---PSFIMGGISQGTGTYLTSNQASQTOETPKPSVGS 1085  
Db 6844 PKVPVPAPVPEIKKKVTEKKVVIK-----KEEAPPKVPPE 6879  
Qy 1086 ISLGLPQOESAKSATLPIYIKOEFSRSPSONSQPGLLVRAQHEGVVVRGTAQIGSGIT 1145  
Db 6880 V-----PKVEE-KRILP--KEEVLVPEVTEPEEPISEEEIPEEPISEEEVAPP 6932  
Qy 1146 RGTPTSKTSVESIPSLRGSIQTGTPA---LPQGTIPTEALVK-GSISMP-IEDSSPE 1198  
Db 6933 RVPEVIKAVPEAPTPVPPKVEAPPKAPKSKKIPKEKVPVQKKKAPPKVPPEKVPPE 6992  
Qy 1199 K-----GREEAA---SKGHVIEGKSGHILSYD-NIKNAREGTSRPTAHSISLRSYSEVE 1251  
Db 6993 KVLVPKKEAVPPAKGRTVLEEKVSAFQEVVVKERLELVVEAEVEEIPPEEEFHEVE 7052  
Qy 1252 GNIKOGMRESPVSAPEGLICRALPRGSHDLKERTVLSGSIQGTTPRATTSTFSDG 1311  
Db 7053 EYFEEG-BFHEVEEFKLE-----QHRVEBEHRVKEVHRVIEVPEAEVEVEFEKP 7101



Db 5628 GKXKPKSEPTVEVUPDEVAQIETVVKDEMIIVEKRRKIKTKRPKSTKVEELFEEQ 5687  
QY 346 -----FPEIRKQOQERFORVQORGAGLSATIAARSEHEISEIIDGLSE 389  
Db 5688 PEEISEPEEVQKEVIEIEIEIEKRLKTKPKLTQOVTETPHE---EII----- 5740  
QY 390 QBNNEKQRLQSVIPPMFADABORVKIINMGLMEDPMKYVKDQFVWVWDHKEIKF 449  
Db 5741 KESEVVOEQEIV-----EKKVKVKV-----KPKTVAERQOL-----KBEIPT 5781  
QY 450 DKFIOHPNFGLIASLYERKSPVDCVLYLYLTKKNENK-----ALVRNYGKRRGRNQOI 505  
Db 5782 EETVE-----EETAEQOOLVVEESKKVKKVKKPTGTVEKTDVDELPGEV--- 5827  
QY 506 ARPSQEEKVEEKDEAKTEKEEKEDEEKDEKEDSKENTKEKIDGTAEETE--- 562  
Db 5828 --PVEVPVEEVEDVAPE--EELIEEQEIVDOIQOQKRVKAKKPKTKIETIEI 5884  
QY 563 -----EREQATPRGRKTKANSQGRKGRITRSMN-----EAAAAA 599  
Db 5885 EEDQPEEEVLOEIIQEQEITEIRQKVKSIKKPKVVTEKTVQDTEQPEESQAEV 5944  
QY 600 AAAATEEPPLPPPP-----EPISPEVETSWTEEMEVAKKGLVEHGRNWAATAK 652  
Db 5945 KETVTEPKKPAPEEAKVEQEKISLKPAPKORLLPEKEQVEEVLKPKKIVAYSE 6004  
QY 653 M---VGTSEAOCKNFYKRRHMLDLQOHKQKTSRKPREEDVDSQCESVASTVSAQ 709  
Db 6005 AEOPETPETEVKEFAIT-----TTEDILDVTKRVKVKKPKTK-----VAEE 6048  
QY 710 EDEDTEAENEENPD--SEVAVKPSDESNSATSRGNTPEAV--ELETTETAPTSPSL 767  
Db 6049 ESTERPABETEFEATQPEVQVEEIP-----EPQVKEVADERKTA----- 6093  
QY 768 AVPSKTKRADESVEVQVNDVSI SAETAQMDVDQOEHSAEESGVCDPPPATKADSDVDEVR 827  
Db 6094 --PKPKPKKEELIEKVEEVALKRVTRPKKEL--PQEAETEEVRL-----KPTQRTSIKPEVK 6147  
QY 828 VPE---NHASKVEGONTKERDLDRASEKV--EPRDSDLVVAQOINNAORPEQSDNDSSATC 883  
Db 6148 LEEVDLQHVKEKEDIQOE--KRTRKVKPKKHEDL-----PEIPDAEPQL 6193  
QY 884 SADEVGDEPERQRMFMDSPKSLNPTGSLVLSPLKPNPLD-----LP 928  
Db 6194 EEAHEIELEKQPK---PEEDQVQPKRGE-----KKQVPEEVLEEKWPSGKRRRLP 6243  
QY 929 QLOHRAAIVPMVSWCTPCNIGTPVSGYALYQRIKAMHSALLESORQOQIDLECR 988  
Db 6244 EQQPEEVQLKPIPS-----KPIEE---QOKPEKAIPQPLVPEEKPESEEELE-- 6289  
QY 989 SSTSPCGTSKSPNREWEVLOPAPHQLITNLPEGVRLPTTRPRPPPLIPSSKTTVASEK 1048  
Db 6290 -----LEP-----LKLPEDK--KPEKPAKKEK-----KKK 6313  
QY 1049 PSFINGGISOGTGTLYTSHNQ-----SYTOETPKPSVSGISLG---LPROQE 1095  
Db 6314 P-----KLKATPSVDEVEAEPPDEPTAEDEVEEMPVDDVKVAVSEDLVLP--EEE 6366  
QY 1096 SAKSATLPIKQEEFSPSONSQPGLLVRAHQEGVVRGTAGATQEGSITRGTPSKISV 1155  
Db 6367 VVPTTEETPEAKOKAHKRTKR-----LKEASVQEQOLLEAAIAE-----I 6407  
QY 1156 ESIPSLRGSITQGNPAL-----PQGIPTALVKSGIS----- 1188  
Db 6408 EKVDEISEIISOKTITLLKKTEDTRPQF--ITYEQLIEDVDRDLKMKVTSNIKEK 6466  
QY 1189 -RMPEDSSPEKGREAAKGHVIEY-----KSGHILSYDNIKNAREGTRSPRTAHEIS 1242  
Db 6467 RRVVLDDSOPLLELITQKR--IQEGDKVADDELIEDQQLIQNQOETTTSEVIGQERK 6524  
QY 1243 L--KRSVESVEGNTKQGMRESVSPAPLEGICRALPRGSPHDLKERTVLSGSIMOGT 1300  
Db 6525 LVKKKKKIKPPIREKLRPQ---CVPEEPTVLCEKVEGVPPFPEIK----- 6568

QY 1301 PRATTESFEDGL-----KYPKOIKRESPPIRAFEGAITKGPYD--GITTIKEMGRSIIHEI 1354  
Db 6569 -----WYFNDILLFASEKYEIVMEQVAKL-----IAKVTSDGVVYCEAKNEAGVAT 6618  
QY 1355 PRODILTQESRKTPEVOSTRPIIEGSIQGPPIKFDNNSGSAIKHNKVSIIITG--PS--- 1411  
Db 6619 SRTNIIILEKQCVPP--QFTKPL-----KIFIEEKQPERLKVTVCQVTKPNPE 6667  
QY 1412 -KLSRGMPLEIVP--ENIKVVERGKVEDKAGETVRSRHTSVSSGSPSVLRTSLHEAPKA 1469  
Db 6668 VKWYKIE--EVIPSETVOM-----FYDEKTDGVAL-----EVINPTPNEAV--- 6707  
QY 1470 QLSPGIYDDTSARRTPVSYQNTMSRGSPPMMNRTSDVTIPPKNSTNHERKSTLTPTQRESI 1529  
Db 6708 ----VYS-----VQAQNGFRAIGNANILSRVDEVPREL---KAPTVTPLSAVVV 6751  
QY 1530 PAKSPVGVDPVVSHPDPHRRGSTAGEVYWSHLPTQLDPAHPHRAALDPAALAAAYLFOR 1589  
Db 6752 PTGGTL-----FFEAKYDGLPRPEVKWRNGREI---IENEETIETETTTTIIK 6798  
QY 1590 QLSPTPGYPSOXYLA-----MENTROTILANDYITSOOMOVNLRPDVARGLSPR----- 1638  
Db 6799 VVNMTRKRTGTKEVWAKNKVGEAKSGSVVSDQKPDQIK---PPRFTQPLEPKYFEGH 6855  
QY 1639 -----EQPLGL-----PYPATR--GIIDLTMNPPTILVPHPGGTSTTPMDRI 1678  
Db 6856 EVAIEAIVESEPLSOFQWVHNEPIKSNEVRIYSQAN--KSTLLIENQSKFVGP--- 6910  
QY 1679 TVIPGTQITFPFPPYNSASMPGHPHTLAAAASABERERERERERERERERERERERERER 1738  
Db 6911 -----FTCRAENVG---GSVTSTATNLIPOEAEFEESPFRVEEL-----V 6949  
QY 1739 RPSGEPQGR-----PGSHGVRSPPSV---RTQETMLQORPSVFOGTNGTSTVITPL 1787  
Db 6950 QVVEVMDGEALLLTQVGT-----KPTPKVEYHNAEKITENKE-----TTISQDL 6995  
QY 1788 DPTAQLRIMPLPAGGPSISQGLPASRYNTAADALAALVDAASAPQMDVSKTESKHEAA 1847  
Db 6996 QGVQCQLOITEVPEPENEGQYECVATNKGKSVKTNVVKQAFYIPDSEITGLTGS----- 7050  
QY 1848 RLEENLRSRSAVBSQ-----QOLEQKTLVEKRSVQCLYTSSAPPSPKQPQ--- 1894  
Db 7051 --EEDLLDRTLSIDQAPKIIKKLPEKIEPKEGEQAKLEVVV-----CKPKPKVKW 7100  
QY 1895 -----HSSVVSAGKDKGPPP---KSRYEELRTRGKTTITTAANFTDVIIT----- 1938  
Db 7101 LRDDQIIFASEEYQIENFEDGTSVLVINRVYPPDL---GTISFEAYNPLGVAVTTALFAV 7157  
QY 1939 RQIASDKDAR--ERGSQSDSSSLSSHRYETPDSAIIV-----ISPASP 1982  
Db 7158 EGIVGSKDYRKPETVWSQEMOVALKAAC--SPSLLNEMDRCAALGETAKESIQFAGNP 7216  
QY 1983 APPQEKLOTYQEV--VKANAQENDPTROYEPLHHYRPOQESPPQQQLPPSSQAEQMGQ 2041  
Db 7217 IP---DIQWYFNWOLRASEKRVWVVOEATLEMK----- 7250  
QY 2042 VPRTHRLTLAD---HICQII-----TODFARNVQSOTPOQPTSTFONSALYSTPV 2093  
Db 7251 -----ITSEDCGYNCKLINEIGMTTRAKFDISS-----TSTIVEETKAKTVKK 7296  
QY 2094 RT--KTSNRYSPESAOQVHHQPGSRVSPENLVKSRGSRPKSPERSHVSSEPEYEPS 2151  
Db 7297 KSGKTKWKRSGASSQNV--QKTEIRIIPTSVAVTSMNVKVKQP----- 7340  
QY 2152 PPQVPPVHEKOD--SLDLLSQRGAEPAEQRN 2180  
Db 7341 ---VSVLVEKSEISVLVKDREVADEERS 7368  
RESULT 31  
Q17490  
ID Q17490 PRELIMINARY; PRT; 5170 AA.





QY 1329 EGATKGPYDGIITIKEMGRS---IHEIPRODILTQSRKTPVQSTRPIIE----- 1379  
Db 2130 -----SGEAFSQFTSEKQDRSDSPHIS-QKEDISQFQNESQEDVKGEQHPDEKPDLE 2183  
QY 1380 --GTSQGTPIKFDNNSQSAIKHNKSLING--PSKLSRGMPLEIYPENI----- 1427  
Db 2184 RQGSYSSGYSKSPGGS-----ITGLDEEAKLSGVQPEPDRPFAESHEKTE 2231  
QY 1428 -----KVVERGKYEDVAKAGETVRSRHTS--VVSSGPSVLRLTLEAPKAQLSPGIYDTS 1480  
Db 2232 ATSDENLFESDKY---APASVPSESDSNRVETITTTTITVREHFEPE-----DHHS 2280  
QY 1481 ARRTPVSVQNTMSRGSPMMNRSTSDVTIPPNKSTNHERKSTLTPT-QRESIPAKSPVPGVD 1539  
Db 2281 ---YVESQEYSSSGSP-----VPSEKSVDRVETITTTTITVREHFEAEDEIPTI- 2327  
QY 1540 PVSHPSPDPHRRSTAGEVYWSHLPTQLDPAMPFHRALDPAALAAALFQRLSPTPGVP 1599  
Db 2328 -----VESH-----DDQAA-----SPVSEED 2345  
QY 1600 QYQLYAMENTROTIL-----NDYITSOOMQVNLRPDVARGLSPREQLPLGLPYPATRGII 1653  
Db 2346 VHEQIQTITTTTITVKEHFVDDDEIDSEHMN-----ESDKYASGSPVSEEDSS 2394  
QY 1654 DLTNMPPILVP-----HPGGTSTPMDRITYPGTQITFPFPPRYNSASMSPGHPTHLAAA 1709  
Db 2395 RVETITTTTITVREHFEDEDDHSPVVOEQY-----SASESP-----VPSE 2435  
QY 1710 ASAE-----ERERERERERERERIAAASDLYLRPGSEQP-- 1745  
Db 2436 KSERVIEITTTTITVREHFEDEHILGQGESDSQIPSESITSENMD---RETSSSPVQ 2492  
QY 1746 -----GRPGSHGYVRSPSPSVRT-----QETMLQORPSVFOGTN 1779  
Db 2493 SNRDEEFVLPAIAPYKQTEYGRVDSDHAPASAPASBPSTQAYKQESQEAHLEQFQ 2552  
QY 1780 GTSVITPLDPTAQL-----RIMPLPAG-CPSISQ----- 1807  
Db 2553 RSVSHEDSPAAYFHONESDHDSPVSDRPLLTQEQHQHQPESGEDGFGSKVLGFA 2612  
QY 1808 ---GLPASRYWTAADALAALVDAASAPQMDVSKTESKHEAARLEENLRSASAVSEQ 1864  
Db 2613 KKAGWAGVVAAPVALAAGAKAA---YDALKKDDDEDPENDEQEKLLPKS---PERQ 2665  
QY 1865 QL-----EQKLEVKRSVQCLYTSTAFSPGKPOPHSSVVYSEAGKDGPPPKSRVEELR 1920  
Db 2666 VLVNVPESDSEISE--ILEYTS-----PSTKSESQCY 2699  
QY 1921 TRGKTITAAFIIDVITRQIASDKDAREGSSQSSDSSSLSHRYETPDAIEVSPAS 1980  
Db 2700 TETVTRTT-----VTREYLDQ-----PQSVTRSRSPSEH-----DISEQYAPES 2738  
QY 1981 -----SPAPQEE-KLQYQYQYVVKANCAENDPTQYEGPLHHY--- 2017  
Db 2739 PVEQDPYVVEKTTVIRQYHDEPPQEEIEQTIPEVTVLREVYESP--EGDEPEQHIET 2796  
QY 2018 -----RPOQESPS----- 2025  
Db 2797 KTTTTITKEVHPVEDVQISPVHSETSVSEKQLPADEQLDEPVTESITATVTRVEE 2856  
QY 2026 PQQLPSPS-SQAEQMGQVPRTHRLITLADHCQI-ITQDFARNQVSSQTPOQPTSTFQN 2083  
Db 2857 PEEVRPPSGSADDESHAPKYMETTTTTTITVREYEVSEDEHQQSQVQORDSPAPSEDS 2916  
QY 2084 SPSALVSTPVTKTSNRYSPESQAVS-----HHORPCSRVSPE 2122  
Db 2917 VKHVIETTTTTTTEETPEDESHSPVSEDDVHGFKVTTTTTTTITVTHEPEPEPPSPDE 2976  
QY 2123 NLVDKSRGSRPKSPERSHVSPEYEPISPPQVPVWH-EKODSLLLLSQGAEPAEARND 2181  
Db 2977 HVVESERVASGSPVSEED--SSRIETTTTTTITVREHFELED-----DQEHVESQEVSA 3030  
QY 2182 ARSPGSIYLP-----FFTKLENTSPMVYKSKQEIFRKLNSGGGSDMAAAQFCTE 2234

Query Match

3.5%; Score 438; DB 5; Length 6994;

Db 3031 SGSP-----VPSEKSERVIEITTTTITVREHFEHDDIPTIVTSHDDPAASSVPSEED 3085  
QY 2235 IFNLPAVTTSGSVSRGSHFADPASNLGLEIIRIKALMGSPDKVEDHGVVMSQPMGVVP 2294  
Db 3086 VHGOIQTTTTTTTTTITVREHVPVDEIDSGRMDLEKYSSESVPSEEDSSRVIE----- 3138  
QY 2295 GTANTSVTSGTREGDPSPHSGVCKPKKLISKSRSKSPKIPGQ----- 2342  
Db 3139 -TTTTTIVS-----REHFEPEDDHSHVSESQEYSASG-----SPVPSEKSERVIEITTTT 3187  
QY 2343 -----GYLGTET-----PSSVSVSHSEG--DYH--RQTPGWAMEDR----- 2374  
Db 3188 TTVTREHFDTEEDYIPSESRTSHDDGITDQHVPSQSPVSEEDDHVTHEQIIRDDPIDEQ 3247  
QY 2375 -----PSSGTSTOFFYNPLTMRMLSTPTPTPIACAPSANQAAPHQONRI-- 2419  
Db 3248 IVESHEYKSESIPEQDSTH-----VIETVMSTPITSERDYDEVEKDVVESADDEIDS 3301  
QY 2420 -----WEREPAPLLSAQYETLSDSD 2440  
Db 3302 STAQYKSESPVQTEKSLLLAKQQQESGDES 3333

## RESULT 32

Q17343  
ID Q17343 PRELIMINARY; PRT; 6994 AA.  
AC Q17343;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE UNC-44 ANKYRINS.  
GN UNC-44.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N2;  
RX MEDLINE=95263663; PubMed=7744957;  
RA Otsuka A.J., Franco R., Yang B., Shim K., Tang L.Z., Zhang Y.,  
RA Boontrakulpoontawe P., Jeyaparakash A., Hedgecock E., Wheaton V.I.,  
RA Sobery A.;  
RT "An ankyrin-related gene (unc-44) is necessary for proper axonal  
RT guidance in Caenorhabditis elegans.";  
RL J. Cell Biol. 129:1081-1092(1995).  
RN [2]  
RP REVISIONS, AND SEQUENCE OF 6126-6994 FROM N.A.  
RC STRAIN=N2;  
RA Otsuka A.J.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U39847; AAB41827.1; -;  
DR EMBL; U21733; AAB38384.1; -;  
DR HSSP; P42773; IINH.  
DR InterPro; IPR000488; -;  
DR InterPro; IPR000906; -;  
DR InterPro; IPR001360; -;  
DR InterPro; IPR002110; -;  
DR InterPro; IPR002383; -;  
DR Pfam; PF00023; ank; 24.  
DR Pfam; PF00531; death; 1.  
DR Pfam; PF00791; ZU5; 1.  
DR PRINTS; PR00001; GLABLOOD.  
DR PROSITE; PS50088; ANK\_REPEAT; 22.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS50017; DEATH\_DOMAIN; 1.  
DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_F1\_1; UNKNOWN\_2.  
DR SMART; SM00005; DEATH; 1.  
SQ SEQUENCE 6994 AA; 775364 MW; 90C4449925D9923D CRC64;

Best Local Similarity 18.7%; Pred. No. 1.7e-14;  
Matches 556; Conservative 356; Mismatches 1102; Indels 958; Gaps 126;

|    |      |                                                               |                         |      |
|----|------|---------------------------------------------------------------|-------------------------|------|
| Qy | 14   | STEQSRYPHPSVOYTFPNTRHOQEF-----                                | AVDPYRSSHLEVSQASOLLQOQ  | 60   |
| Db | 2599 | SSEQOSEPHIVKETTTTTTTTRELVDPEKGNVTFPAPSSHAESERQVPSPVVSQHE      | 2658                    |      |
| Qy | 61   | QQQ--QLRRRPSLLSEFH-----PGSRP-----                             | QERTSYEPHP--GPSVDHDSLES | 106  |
| Db | 2659 | YPVVVETTTTNTVTSNIYDDNVPSSDPATQHQQOSETSVHRSHPDPSVEESDGEGLS     | 2718                    |      |
| Qy | 107  | KRPRLEQVSDSHFORVSAVPLVHPLPEGLRASADA--KQDPAGFGKHEAPSSPI-----   | 161                     |      |
| Db | 2719 | KVLGFAKKAGVAGVGAAPVALA-----AVGAKAAYDAFEKDEDDDETSHSPSPVPEYQS   | 2775                    |      |
| Qy | 162  | GQPCGDDQNAS-----PSKLSKELIQSMQDVRDREIAKVEOOILKLKKKQOOLBEE      | 212                     |      |
| Db | 2776 | BOYQDDSAQSSHTDEHHMPSEPIHEKETEEFDH-----                        | SHPSVLSKERDDQVTSSE      | 2830 |
| Qy | 213  | AA-----KPPEPK-----PVSPPP-----                                 | VEOKHRSIVQII-----       | 239  |
| Db | 2831 | TTTTTTTVEYNDEPEDQEHQHPSPAPSSHTAEAPHIVETTTTTTTTREFQEEPEEL      | 2890                    |      |
| Qy | 240  | ---YDENRKKAEBAHKIFEGLGPKVELPLYNOPSOTKYVHENIKTNQVMRKKLILFFKRR  | 296                     |      |
| Db | 2891 | EYKQDNRKSPSSH-----SOENLVTTETTTTTTTTSE-----                    | YDEP                    | 2928 |
| Qy | 297  | NHARKORQKICQRYDQLMAEWKKV---DRIENPNRRKAKESKREYVEYKOPPELRQRE    | 354                     |      |
| Db | 2929 | EHFEQETGKSPAPSSHVES--EKQVLSPVASDPRH--VMETTTTTTTTTRQFHDDDESRS  | 2986                    |      |
| Qy | 355  | QOERQVRQGRGAGLSATIASEHISEIIDG-----LSEOE--                     | NNEKOMRQLSVIP           | 404  |
| Db | 2987 | DSPNRDVV--EAQSIHSSNIESHQOFSEKDDDSQSGPSPVSEDEPVKHEGKYKFTETITE  | 3044                    |      |
| Qy | 405  | PMFEDAQRVKFINNGLME-----DPMKYKDRQPMNVTMDHEKEIFKDKFIQHPKNF--    | 459                     |      |
| Db | 3045 | TREDFRPE-----LDSERLSEPAQSPMEV-----                            | SPITTEENIVRPSSSVK       | 3085 |
| Qy | 460  | -----GLIASYLERKSVPCVLYLYLTKNENYKALVRYNKGKRGGRNQOIAQPSQEEK     | 513                     |      |
| Db | 3086 | SEYSGEHPVSVETKTTTTTTRFYDDQDE-----                             | HENQSQSEELSSIPTEEEE     | 3137 |
| Qy | 514  | VEEK---BEDKAETKKE--BEKKDEEEKEDSKENTREKDKIDGTABETEEREQAT       | 568                     |      |
| Db | 3138 | HEGSHLFKETTITTTTTRFYDEPENVEELQDPQSPAPSSHVESGIIHASESPVAQQOEI   | 3197                    |      |
| Qy | 569  | PRGRKTANSQGRKGRITRSMTNEAAA-----SAAAAATEEPPLPPPPPISTEP         | 622                     |      |
| Db | 3198 | PQ-----TREFHEDSPAAQYFHEEYESHVLTEQAPLLTEOQHQPESGDE             | 3242                    |      |
| Qy | 623  | VETSRWTEEMEVAKK--GLVEHGRNWAIA--KMVTKSEAOCKNFYFNKRRHNLNLLQ     | 680                     |      |
| Db | 3243 | SDGELGSKVLGPAKKAGVAGVVAAPVALAAVAKAAYDALKKDDDEEDQEERESLLR      | 3302                    |      |
| Qy | 681  | QHKQ-----KTSRPREEDVSCQESVASTVSAQEDE-----DIEA                  | 716                     |      |
| Db | 3303 | QERSIDSPHASEQSQIEEHEHFEESPVPSEKHVHTTTTTTTTTTTRFYFDEHEPLVSQEI  | 3362                    |      |
| Qy | 717  | SNEENPDSE-----VEAVKPSDESPENAT--SRGNTERPAVE                    | 752                     |      |
| Db | 3363 | BEGKGSADSKSLPHVVVETTTTTTTTTRFYDKNDSESPVPSEKEQEPTTVSREYVTAEG   | 3422                    |      |
| Qy | 753  | LEPT---TETAPST-----SPSLAVPSTKPA-----                          | EDESVE                  | 781  |
| Db | 3423 | DEPHHYTETTTTTTVEYKEVDQEMGDDDDPEPKQESQVETTTTTTTTMSRYDNDNDETR   | 3482                    |      |
| Qy | 782  | TQVNDISIAETAQMDV-----DQOHSABEGSVCDPPP-----                    | ATKADSDVY               | 824  |
| Db | 3483 | SEAGDSHITETTKTTTVVREFHGEQPEETEETDEVEELPPKIEEDNVDNVEYSESSITSYR | 3542                    |      |
| Qy | 825  | EVRYPENHASKVEGDNTERDLDRASEKVEPRDEDLVVAQINAQRPQSDNDSSATCS      | 884                     |      |

|    |      |                                                               |      |
|----|------|---------------------------------------------------------------|------|
| Db | 3543 | EVRPDEPHIETTTTTTTTTRFYHNEPEETYDDQKD---AAPISFS--QEHQDDDSQAS--  | 3595 |
| QY | 885  | ADEDVGEPEQRMPDMDSKPSLLNPTGTSILVSSPLK-----PNPLDLPOLQH          | 932  |
| Db | 3596 | -----HDQHDRESPVSEKSVKHTTTTTTTTTRLYDYDEASEIRGESPVATEEH         | 3648 |
| QY | 933  | RAAVIPMVCSTPCNICPIGTVPVGYALYQRIKAMHESALL-----EBORQROEQ        | 982  |
| Db | 3649 | -----VSSTKSDSE-----SEQHVPVSIETTTTTTTTTRFYDDQDELQREDH          | 3690 |
| QY | 983  | IDLECRSTSPCGTSKSPNREWEVLQAPHQLIINLPGVRLPTTRTR---PPPLPLPS      | 1039 |
| Db | 3691 | TQSEERSIP--TEETHEDSHLIKET-----TTTTTTRFYDEPENVEK               | 3735 |
| QY | 1040 | SKTTVASEKPSFIMGGSISQGTPTGYLTSNQAS--YTQETPKPSVGSIISLGLPQOESAK  | 1098 |
| Db | 3736 | LQDSQFSLSPS-----SHVESEIYPESP-----VAKQOEIPQ                    | 3768 |
| QY | 1099 | -----SATLPYIKOEFSRPSQNSQPEGLLVRAHQHGVVGRGTAGAIQEGSITRGTPTS    | 1151 |
| Db | 3769 | TREFHEDSPAQYFHEDEYEHQVPTQAP--LLTEQHQHP-----ESGEESDGEFG        | 3818 |
| QY | 1152 | KISVESIPSLRGSITQCTPALPQGTIPTEAL--VKGSIISRMPIEDSPSEKGREEAASKGV | 1210 |
| Db | 3819 | S--KVLGFAKKAGVAGVGAAP---VALAAGAKAAYDALKKEDD--EEDQERES---L     | 3869 |
| QY | 1211 | IYEGSGHILSYDNINKNAREGTRSPRTAHEISLARSYESVEGNIKQGMMSRSPVASPLE   | 1270 |
| Db | 3870 | IREER-----SFD-----SPHASEQSQIEKEHRFE-----ESPVPSEKH             | 3904 |
| QY | 1271 | GLICRALRPGSPHSDLKE--RTVLSSGIMQCTPRATTFESFDGLKYPKQIKRESPPIRAF  | 1328 |
| Db | 3905 | HDQSSALPOESVSPKESRFTNDESEFGVKSQDHYTEDQESLKSPE-----            | 3953 |
| QY | 1329 | EGATKPKPYDGIITIKEMGRS---IHEIPRODILTOESRKTPEVQVOSTRPIE-----    | 1379 |
| Db | 3954 | -----SGEAFSQFTSEKQDRSDSPIHS--QKEDISQFQNESSEQPHDEKPDLE         | 4007 |
| QY | 1380 | --GSISQETPIKFDNNGSQAIKHNKSLTNG--PSKLSRGMPPLEIVPENI-----       | 1427 |
| Db | 4008 | ROGSYSSGYSPKSGGS-----ITGLDEEKALSQVQEPEDRPFNFAESHEKTE          | 4055 |
| QY | 1428 | -----KVVERGKYEDVKAGETVRSRHTS--VYSSGSPVLSRTLHEAPKAQLSPGIYDDTS  | 1480 |
| Db | 4056 | ATSDENLFESDKY---APASVPVEDSSNRVETTTTTTTTTRFHEFEPE-----DDHS     | 4104 |
| QY | 1481 | ARRTPVSYQNTMSRGSPMMNRTSDVTTPPNKSTNHERKSLTPT--QRESIPAKSPVPGVD  | 1539 |
| Db | 4105 | ---YVWESQEYSSSGSP-----VPSEKSDRVETTTTTTTTTRFHEFAEDEIPTI--      | 4151 |
| QY | 1540 | PVSHSPDPHHRGSTAGEVYWSHLPTQDLPAMPFPHRALDPAAYALFQRLSPTPGYPS     | 1599 |
| Db | 4152 | -----VSSH-----DQAA-----SPVSEED                                | 4169 |
| QY | 1600 | QYQLYAMENTROTIL-----NDYITSOQMOVNLRPDVARGLSPREOPLGLYPATRGII    | 1653 |
| Db | 4170 | VHEQIQTITTTTTTTTKEHFVDPDEIDSEHM-----ESDKVAGSPVPSEEDS          | 4218 |
| QY | 1654 | DLTNMPTILVP---HPGGTSTPMDRITYPGTQITFFPPRPNYSASMSGPHTHLAAA      | 1709 |
| Db | 4219 | RVTETTTTTTTTTRFHEFEDDHPVWQTOEY-----SASEP-----VPSE             | 4259 |
| QY | 1710 | ASAEK-----ERERERERERERERERERERERERERERERERERERERERERERER      | 1745 |
| Db | 4260 | KSERVETTTTTTTTTRFHEFDEHILGQOQESDSQIPSESITSENMD---RETSSPVQ     | 4316 |
| QY | 1746 | -----GRFGSHGYVRSPSPSVRT-----QETMLQQRPSVFGTN                   | 1779 |
| Db | 4317 | SNRDEEVLPAIAPYKQTEYGRVDSHDAPASPAPSAESPIQAYKQESQEAHSLQFQ       | 4376 |
| QY | 1780 | GTSVITPLDPTAQL-----RIMPLPAG--GPSISQ-----                      | 1807 |
| Db | 4377 | RSSVSHEDSPAQYFHDNESHDSHSPVSPDRAPLLTEOQHQPESGEESDGEFGSKVLGFA   | 4436 |

|                                   |                                                                       |                                                              |                                                        |                        |                       |
|-----------------------------------|-----------------------------------------------------------------------|--------------------------------------------------------------|--------------------------------------------------------|------------------------|-----------------------|
| QY                                | 1808                                                                  | ---                                                          | GLPASRYNTAADLAALVDAASAPOMOVSKTKESKHEAARLEENLRRSAAVSEQQ | 1864                   |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 4437                                                                  | KKAGTWAGGVVAAAPALAAVAGAKA---                                 | YDALKKDDDEDENPDQEKLLPKS---                             | PERQ 4489              |                       |
| QY                                | 1865                                                                  | QL---                                                        | EOKTLEVKRSVOCLTSSAFFSGRQPQHSSVVVYSEACKDGKPPKPSRYEELR   | 1920                   |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 4490                                                                  | VLNVNVPESDSEISE---                                           | ILEYTS---                                              | PSTKESQCY 4523         |                       |
| QY                                | 1921                                                                  | TRGKTTTAAANFIDVITIRQIASDKDARERGSQSDSSSSLSHRYETTPSDAIEVSPAS   | 1980                                                   |                        |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 4524                                                                  | TETVRTTT---                                                  | VTREYLDD---                                            | PQSVYTRSRSPSEH---      | DISQYAPES 4562        |
| QY                                | 1981                                                                  | -----                                                        | SPAPQOE-KLQTYQPVVVKANAQENDPTRQVEGPLHHY---              | 2017                   |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 4563                                                                  | PVEQDPYVVEKTTTIVIRQYHDEPPOEIBEOTIPEEVTVLVREYESP--            | EGDEPEQHYIET                                           | 4620                   |                       |
| QY                                | 2018                                                                  | -----                                                        | RPOQESPS-----                                          | 2025                   |                       |
| Db                                | 4621                                                                  | KTTTTTKEVHVPEEDVOISPVHSETSVSEKQLPADQLDEPVTESATTATVTRERYEE    | 4680                                                   |                        |                       |
| QY                                | 2026                                                                  | PQOOLPPS-SOAEQMGQVPRHLRLITLADHICQI--                         | ITQDFARNQVSSOTPOQPPSTFQON                              | 2083                   |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 4681                                                                  | PEEVPPSGSEADDESHAPKYMETTTTTTTTREVYEVSDEDEHQOSQVQRDESPAPSEDS  | 4740                                                   |                        |                       |
| QY                                | 2084                                                                  | SPSALVSTPVRYKTKSNRYSPEQAQSV-----                             | HQRPGRSVSPSE                                           | 2122                   |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 4741                                                                  | VKHVIEKTTTTTTTBERYEPEDSHSPVSEDDVHGCFVKTTTTTTTTHVHEFEPEDPPSDE | 4800                                                   |                        |                       |
| QY                                | 2123                                                                  | NLYDKSRGRGPKGSPERSHVSSEPEPTISPPQPVVH--                       | EKQDSLILLSQGAEPAEORND                                  | 2181                   |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 4801                                                                  | HVVESERYASGSPVPEED--SSREIETTTTTTTTTRHFELED-----              | DQEHVYESQEYSA                                          | 4854                   |                       |
| QY                                | 2182                                                                  | ARSPGSIYLPSS-----                                            | FFTKLENTSPMVKKSKQEIFRKLNSGGSDMAAQPGE                   | 2234                   |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 4855                                                                  | SGSP-----                                                    | VPSEKSERVIETTTTNTVTRHFEHEDDIPVITVSHDDPAASSVPSEED       | 4909                   |                       |
| QY                                | 2235                                                                  | IFNLPAVTTSGSVSRGSHGFADPNALGLEDIIRKALMGSDDKVEDGHVVMQPMGVVP    | 2294                                                   |                        |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 4910                                                                  | VHGQIQTTTTTTTTRHVHVPEEDISGRMDELEKYSSESVPVPEESDSSRVIE         | 4962                                                   |                        |                       |
| QY                                | 2295                                                                  | GTANTSVTSGTREEDGDPSPHSGCVCKPKLISKSNRKSKSPIQG-----            | 2342                                                   |                        |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 4963                                                                  | -TTTTTTS-----                                                | REHFEPEDDHSHVYESQYFASG-----                            | SPVPSEKSERVIETTTT 5011 |                       |
| QY                                | 2343                                                                  | -----                                                        | CYLGTGER-----                                          | PSSVSSVHSG--DYH--      | ROTQGWAMEDR----- 2374 |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 5012                                                                  | TTVTRHFEFDTEEDYTPSSRSTSHDDGIDTDQHVQSPVPESEDDHVTHEQIIRDDPIDEQ | 5071                                                   |                        |                       |
| QY                                | 2375                                                                  | -----                                                        | PSTGSTQPPYNPLTMRLSSTPTPTPIACAPSAYNQAAPHQONRI--         | 2419                   |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 5072                                                                  | IVESHEYKSESSIPSEODSTH-----                                   | VIEVMTSTPITSERYDPEVEKDVVESADDEIDS                      | 5125                   |                       |
| QY                                | 2420                                                                  | -----                                                        | WEREPAPLLSQAQYETLSDSD 2440                             |                        |                       |
|                                   |                                                                       | : :                                                          | : :                                                    | : :                    |                       |
| Db                                | 5126                                                                  | STAQYKSESPPVQTEKSLLLAKQOQESGDES 5157                         |                                                        |                        |                       |
| RESULT 33                         |                                                                       |                                                              |                                                        |                        |                       |
| Q9NHX6 PRELIMINARY; PRT; 1966 AA. |                                                                       |                                                              |                                                        |                        |                       |
| ID                                | Q9NHX6                                                                |                                                              |                                                        |                        |                       |
| AC                                | Q9NHX6;                                                               |                                                              |                                                        |                        |                       |
| DT                                | 01-OCT-2000 (TReMBLrel. 15, Created)                                  |                                                              |                                                        |                        |                       |
| DT                                | 01-OCT-2000 (TReMBLrel. 15, Last sequence update)                     |                                                              |                                                        |                        |                       |
| DT                                | 01-MAR-2001 (TReMBLrel. 16, Last annotation update)                   |                                                              |                                                        |                        |                       |
| DE                                | GRUNGE.                                                               |                                                              |                                                        |                        |                       |
| GN                                | GUG OR CG6964.                                                        |                                                              |                                                        |                        |                       |
| OS                                | Drosophila melanogaster (Fruit fly).                                  |                                                              |                                                        |                        |                       |
| OC                                | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;         |                                                              |                                                        |                        |                       |
| OC                                | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; |                                                              |                                                        |                        |                       |
| OC                                | Ephydroidea; Drosophilidae; Drosophila.                               |                                                              |                                                        |                        |                       |
| NCBI                              | Taxid=7227;                                                           |                                                              |                                                        |                        |                       |
| OR                                | [1]                                                                   |                                                              |                                                        |                        |                       |

| RP | SEQUENCE FROM N.A.                                                    | Query Match           | 3.4%;             | Score 433.5;       | DB 5;       | Length 1966; |
|----|-----------------------------------------------------------------------|-----------------------|-------------------|--------------------|-------------|--------------|
| RA | Erkner A., Roure A., Core N., Angelats C., Vola C., Fasano L.,        | Best Local Similarity | 19.6%;            | Pred. No. 5.3e-15; |             |              |
| RA | Keridge S.;                                                           | Matches 421;          | Conservative 243; | Mismatches 770;    | Indels 709; | Gaps         |
| RT | "Grange is required for proximal and ventral leg development in       |                       |                   |                    |             |              |
| RT | Drosophila.";                                                         |                       |                   |                    |             |              |
| RL | Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.               |                       |                   |                    |             |              |
| DR | EMBL; AF217844; AAF34752.1; -.                                        |                       |                   |                    |             |              |
| DR | FlyBase; FBgn020427; Gug.                                             |                       |                   |                    |             |              |
| SQ | SEQUENCE 1966 AA; 208033 MW; 96AF90E2082E770C CRC64;                  |                       |                   |                    |             |              |
| QY | 385 DGLSEQENNEKOMRQLSVIPPMFADAEORRVKFINNGLIMEDPMKYKDRQFMNVTDHE 444    |                       |                   |                    |             |              |
| Db | 82 DGLAASRDDTPINALDLVHDSYDPGKA-----LQALVKCPVSGKIDKK-----WTSGE 131     |                       |                   |                    |             |              |
| QY | 445 KEIFKDKFIQHPKNFGLI-ASYLERKSPDCVLYYYLTKK-----NEN-----YKALVRR 493   |                       |                   |                    |             |              |
| Db | 132 TKKFIKGLRQFGNFKRIHKDLLPHDKDTPELVFVYLMKKTKPCANNRPHRRRQSLAR 191     |                       |                   |                    |             |              |
| QY | 494 NYGKRGRNQOIARPSQEKVE-----EKEEDKAEKTEKKKEEKDEEEK 538               |                       |                   |                    |             |              |
| Db | 192 NRYTRANNSNSNTPPKKEDTPFPQTATTATAATAASETASRSPSPAVSKEENSLSUEDDA 251  |                       |                   |                    |             |              |
| QY | 539 DEKEDSKENTKEKDKIDG---TAEETEEREQATPRKRTANSOGRKGRITRSMTEAAAA 595    |                       |                   |                    |             |              |
| Db | 252 SECDSDSLTHKKDESPSRMRTNRKQNNNSTSGNNTAGNGGNATSISSGTTGGAA 311        |                       |                   |                    |             |              |
| QY | 596 A-----SAAAAAATEEPPPLPPPEP ISTEPTVETSRWTEEMEAVAKGLVEHGRNWA 648     |                       |                   |                    |             |              |
| Db | 312 GGNSSKDKQSANAVANGKRPKRGETPDVSGGASDPSK-----TPTKVAEASSAN-- 362      |                       |                   |                    |             |              |
| QY | 649 AIAKMVGTKSEACKNFYNYKRRHNLNLLQHQKTKSRKPREERDVQSCESVAS----- 704     |                       |                   |                    |             |              |
| Db | 363 ---KRGGKGQDTPNKK-----KRTQESNEPSAHEENAIKERRKRPD-SPVESMNSDSRP 413   |                       |                   |                    |             |              |
| QY | 705 -----TVSAQDEDIEDIASNEEENPEPDESEAVKPSDES 737                       |                       |                   |                    |             |              |
| Db | 414 DSVLDDGESNTTDTTBAQOSTKDSKETVSCKEERKWTNDLEAKAEKAKAEALAEADS 473     |                       |                   |                    |             |              |
| QY | 738 PENATSRGNTPEAVELEPTTETA-----PSTSPS-LAVPSTKP-----AEDESVEQVND 786   |                       |                   |                    |             |              |
| Db | 474 KDSAIKNMDEETNIQAPSSADTSLVDGPNPNALPSPVAAPITMKVPTTIATVEALNASVDR 533 |                       |                   |                    |             |              |
| QY | 787 SISAETAEQMDVD-----QOEHAEBEG-----SVCDDPPPATK 818                   |                       |                   |                    |             |              |
| Db | 534 KEAIEKMECSDSPEMLKKLATIKQEVSPQOQHMQOOSQOQOOLAPVGIPOPPSPCP 593      |                       |                   |                    |             |              |
| QY | 819 ADSVDVEVRVPENHASKVEGDNTKE-RDLDRASEKVEPRDEDLVVAQOINAOBPQSDN 877    |                       |                   |                    |             |              |
| Db | 594 SESVYIKKEPMEDSMATCNQSNPEQDL---KVIEIKNED---ALKHSGAGLPP----- 643    |                       |                   |                    |             |              |
| QY | 878 DSSATC---SAEDVDVGEPERMFP-----MDSRPSLLNPTGSLVSPK----- 921          |                       |                   |                    |             |              |
| Db | 644 --SGCAPPSALHPLTSGAPVESQEBPLHLQHPHGQVTTQPPPGYLIDQLKYGSPGGV 701     |                       |                   |                    |             |              |
| QY | 922 -PNPLDLQLOHRAA---VIPPWVSCTPCNTP-----IGTPVSGVALYQR 962             |                       |                   |                    |             |              |
| Db | 702 PPOF---POLHSDAAGVSGVAPPAPTTPQKYPPPEMEMKFAPODLKYPPPPPLDALKYSQ 758  |                       |                   |                    |             |              |
| QY | 963 HIKAMHESA-----LLEEQRQRQSID----- 984                               |                       |                   |                    |             |              |
| Db | 759 EQQAAAAAACKYDMKYMMQOQKYNVELSAAHQPSKPGYQDSLKIPDKPGFGHLP 818        |                       |                   |                    |             |              |
| QY | 985 -----LECRSSTSPGTSKSPNRWEVLQAPAPHOLITNL---PEGVRLPTT----- 1027      |                       |                   |                    |             |              |
| Db | 819 HNVGSPLDAAHKYGPPTTSQESQQOOP--QPPAHQVPPGATPPGCIAMPKPHYQHDVQTP 876  |                       |                   |                    |             |              |
| QY | 1028 ---RPTRP-----PP-----PLIPSSKTTVAASEKPSPTIMG----- 1055             |                       |                   |                    |             |              |

RESULT 33

09NHX6

ID Q9NHX

AC Q9NHX

DT 01-OC

DT 01-OC

DT 01-MA

DE GRUN-  
GEN

GN  
OS

Eukaryotic

Pteryg

∞ Ephyd:

OX NCBI\_

RN [1]



DR InterPro; IPR003006; -  
DR Pfam; PF00047; ig; 7.  
SQ SEQUENCE 3263 AA; 37

Query Match 3.4%; Score 432.5; DB 5; Length 3263;  
Best Local Similarity 18.7%; Pred. NO. 1.2e-14;  
Matches 472; Conservative 361; Mismatches 905; Indels 791; Gaps 114;

[illegible]

```
Db 2185 QPVEVMDGEALLLTQCVTG-----KPPKPVWYHNAEKITENKE-----TTISQDL 2230
Qy 1798 DPTAQLRIMPLPAGPFSISQGLPASRYNTADALAALVDAASAPQMDVSKTKESKHEAA 1847
Db 2231 QGVCLQITEFPPENEGYECVATNKIGKSVKTNVKTQAFYIPDSEITGLTGS----- 2285
Qy 1848 RLEENLRSRSAVSBQ-----QQLEOKTLEVEKRSVOCLYTSSAFPSPKQOP----- 1894
Db 2286 --EEDLLDRTLSIDQAPKIITKKLEKPEKEGEQAKLEVWV-----GKPKPKVKW 2335
Qy 1895 -----HSSVYISBAGKDKGPPP---KSRVEEELTRGKTTITAAFNIDVITRQIASD 1944
Db 2336 LRDDQIFASEYQLENFEDGTSVLVNHVYPDDL---GTISFEAYNPLGVAVITLFA- 2391
Qy 1945 KDAREGSSQSSSSLSHRYPETSDAIEVISPASSAPPOEKLOTQOPEV-VKANQAE 2003
Db 2392 -----VEGIVGSKDYRKP---EWVSQEMEQQVALKDQIOWTFNNVQLRASEKY 2435
Qy 2004 NDPTRQYEGPLHHYRPOQESPSPQQLPPSSQAGMGQVPRTHRLITLAD---HICQII- 2059
Db 2436 RMVQEQEATLEIMK-----ITSEDCGYYNCKLIN 2465
Qy 2060 ----TQDFARNQVSSQTPQPPSTFQNSPALSALVSTPVRT---KTSNRYSPSPSAQSVHHQ 2113
Db 2466 EIGMTMTRAKFDISS-----TSTIVETRAKTTVKKKGKKTVMKRSGASESNV-Q 2516
Qy 2114 RPSRVSPENLVDRSGRSPKSPRSVSSEPEYPIPPQVPVHVHKQD--SLLLLSQR 2171
Db 2517 KTEIRIIPTSAVETSMNVIKVQKP-----VSVLVEKSEISeVLVWDR 2559
Qy 2172 GAEPAEQRN 2180
Db 2560 EVADAERS 2568

RESULT 35
Q9VSK5
ID Q9VSK5 PRELIMINARY; PRF; 1963 AA.
AC Q9VSK5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE CG5964 PROTEIN.
GUG OR CG5964.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; brachycera; Muscomorphia;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadenot E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.B., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
```

```
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003555; AF50413.1; -
DR Flybase; FBgn0020427; GUG.
DR InterPro; IPR001005; -
DR InterPro; IPR001005; -
DR Pfam; PF00249; myb_DNA-binding; 1.
DR Pfam; PF01448; ELM2; 1.
DR SMART; SM00395; SANT; 1.
SQ SEQUENCE 1963 AA; 207927 MW; 8F386DE56E66CB68 CRC64;
```

Query Match 3.4%; Score 432; DB 5; Length 1963;  
Best Local Similarity 19.0%; Pred. No. 6.4e-15;  
Matches 437; Conservative 250; Mismatches 910; Indels 702; Gaps 100;

```
Qy 382 EIIDGLSEQENNEKQMRQLSVIPPMFD-----AQRRVKFINM--NGLMEDPMKV 431
Db 8 EIRVPGCHQETDERELEESRWSPGVADGDLMLFLRAARSMAAFQGMCDGLEDGLAAS 67
Qy 432 KDRQFMV-----WTDEKEIFKDKFTQHPKNFL 461
Db 68 RDDTTINALDVHDSGYDPGKALQALVKCPVSKGIDKKWTEDETKFKIKGLRQKGNFR 127
Qy 462 I-ASVLERKSPDCVLYLYLTKK-----NEN-----KALVRNRYGKRRNQIARSPQ 510
Db 128 IHKOLLPHKDPTELVEFYLLWKTPGANNRRPHRRRROSALRRNRVTRANNNSNTPPKK 187
Qy 511 EEKVE-----EKEEDKAETKEEKEDEKEDKESKENTREKDKID 555
Db 188 EDTPEPQTATTATAATAASATSRSPAVSKENSSLTEDDASECDSLSLTHKRDESP 247
Qy 556 G---TAETEEREQATPRGRATANSQGRKGRITRSMTNEAAAA-----SAAAAAATE 605
Db 248 SRMTRNKQNNNSSTSGNNTAGNGGNATISGSGTGGAGAGNSSSKQDSANAVANG 307
Qy 606 EPPPLPPPPPISTEPVETSRMTEEMEYAKGLVHEGRNMAAIAKMGVTKSEAKCNF 665
Db 308 KRPRGSETPDVSGASVDSPK-----TPTKAVAESAN-----KRKGKQDTPNKK- 354
Qy 666 YFNKRRHNLNLLQHQKQTSRKPREEDVSQCESVAS----- 704
Db 355 ---KRTEQESNEPSAHEENAIEKKRKPDP-SPEVSMNDSRSPSVLDDGSGNSTDTTTTA 409
Qy 705 -----TVSAQEDDEDIEASNEEENPESEVAVKPSDENATSRGNTPEPAVELE 754
Db 410 EQQSTKDSKEIVSCKEEREMVTNDLEAKAEKAKAEALAEKSDSAIKNMDTEINTQAP 469
Qy 755 PTTTAPSTPSLAVPSTKPAEDSVETQVNDISIAETAQMDVDDQOQSHSAEESVCDPP 814
Db 470 SSAATSLVDGPN---PNAIPSPAAPITKMPVTIATVATVATVATVATVATVATVATVAT 525
Qy 815 PATKADSVDEVVRYPENHASKVEGDNTERKDLDRASEKVE-----PRDELVVAQIINAQ 869
Db 526 PEMLKLATIKQEVSPQQOQHMQOOSQOQQOQLAPGVIPQPPSPCPSPSESYII----- 578
```















Db 2327 APTVEPTVEKLAPEVSKETSEVE-----PAAIVEQKDL----- 2359  
QY 1958 SSSLSSHRYETPSDAIEVISPASS-----PAP-----POEKLQTOYQVPPVVKANQ 2001  
Db 2360 --PVPETSAPTVEPTVEKLAPEVSKETSEVEPAEIVEQKDVVPPTSAPTVEPTVEKLAPE 2417  
QY 2002 AENDPTROYEGLPHYRQOQSPSPQOQLPPSSQAEGMGQVPRHRLITLADHICQIITQ 2061  
Db 2418 VESKETSEVEPA--EIVEQKDVVPPTSAPTVE-----PTVEKLAPEVES----- 2459  
QY 2062 DFARNQVSSQTPQ-----PTSTFQNSPSPALVSTPVRTKTSNRYK----- 2102  
Db 2460 ----KETSEVEPAEIVEQKDVVPPTSAPTVEPTIEKLAPEVSKETSEVEPAEIVEQKDV 2515  
QY 2103 --PESQAQSVH-----HORPGSRVSPENLVDSKSGS-----RPGK 2135  
Db 2516 SVPEPTSAPTVEPTIEKLAPEVSKETSEVEPAEIVEQKDVVPPTSAPTVEPTVEKLAPE 2575  
QY 2136 SPERSHVSEPEYISPPQVPV-----VHEKQDSLILLSQGAEPAE-- 2177  
Db 2576 SKETSEV--EPAEIVEQKDVVPPTSAPTVEPTVEKLAPEVSKETS-----EVQPAEI 2626  
QY 2178 --ORNDARSPGSIYLPSTFKLENTSPMVSKKQEIFRKLNSGGGSDMAAQ----- 2230  
Db 2627 VEHKDVQVPETSS--PTVEPTVEKLAPE--VESK-----TSEVEPAEIVEQKD 2670  
QY 2231 ---PGTEIFNFI--PAYTSGSVSSRGHSPADPASNLGLDI--IRKALMGSEFDKVEDHGCV 2285  
Db 2671 VVPPTSAPTVEPTVEKLAPEVSKETSEVEPAEIVEQKDVVPPTSAPTVEPTVEKLAPEV 2730  
QY 2286 MS-----OPMGV-----VPGTA-----NTSVTSVSGTTRREEDGDS 2315  
Db 2731 ESKETSEVQPAEIVEHVKDVQVPETATTFTPTKEKLAPEVSKETSEVQTAIVEQKDV 2790  
QY 2316 PHSGGV-----CKPKLI-----SKSNRKSPIPGQYGLTERPSS-----VSSVHSE 2359  
Db 2791 PETSAVPTVEPTVEKLAPEVSKETSEVQQAIVEQKDVVPPTSATTVEPTKEKLAPEVSK 2850  
QY 2360 GDYHRQTPGWA--WEDRPSSTGSGTGFYNNPLTMRMLSSPTTPIACAPSAVNOAAPHOQNR 2418  
Db 2851 ETSEIQTAAIVEQKDVVPPTSSTSY-----VEPTKEKLAPEVSKETSEVQQA 2898  
QY 2419 IWREPEAPLLSAQYETLSDSD 2440  
Db 2899 IVEQKDVVPPTSATTVEPTKE 2920  
RESULT 39  
O43161 ID O43161 PRELIMINARY; PRT; 3851 AA.  
AC O43161  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-AUG-1999 (TrEMBLrel. 11, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE BASSOON PROTEIN (K1AA0434) (FRAGMENT).  
GN BSN OR KIA0434.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Winter C., Dieck S.T., Bockmann J., Boeckers T., Kaempf U.,  
RA Sanmarti-Vila L., Langnaese K., Altmann W., Stumm M., Soyke A.,  
RA Wieacker P., Garner C.C., Gundelfinger E.D.;  
RT "The presynaptic cytomatrix protein Bassoon: Sequence and chromosomal  
RT localization of the human BSN gene";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE OF 2281-3851 FROM N.A.  
RP TISSUE=BRAIN;  
RC Ishikawa K., Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N.,  
RA Tanaka A., Kotani H., Nomura N., Ohara O.;

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y18448; CAA77176.1; .  
DR EMBL; Y18449; CAA77176.1; JOINED.  
DR EMBL; Y18450; CAA77176.1; JOINED.  
DR EMBL; Y18451; CAA77176.1; JOINED.  
DR EMBL; AB007894; BAA23707.1; .  
FT NON\_TER 1  
SQ SEQUENCE 3851 AA; 410202 MW; 7CA91DB23284C161 CRC64;

Query Match 3.4%; Score 425.5; DB 4; Length 3851;  
Best Local Similarity 18.8%; Pred. No. 3.6e-14;  
Matches 568; Conservative 355; Mismatches 1132; Indels 967; Gaps 134;

QY 30 PNTRHQEFAYPDYRSSHLEVSQASQLLQOQQOQLRRRPSLLSEFFHPSRDPQERRTSY 89  
Db 458 PPTSQQPPVGAP--HRAS-----GTSPLKQKQPGQGLQPSGFL-----PAKASPLSTRKAS 506  
QY 90 EPFHPGP--SPVDHDSLESKRPRLEQVSDSHFQVRAVLAFLVHPLPE----- 135  
Db 507 LPSKASQAKPL--RASEPSKTPSSVQEKKT--RVPTKAEPMPKPPPTTTPATPKVKS 562  
QY 136 GLRASADAKDPAGGKHEAP-----SSPISQPCGDDQNASPSKLSKEELIOSMDRV--- 188  
Db 563 GVRRAEPA--TPVVKAVPEAPKGGAEADVKGYSQDASRSPQSL--DTGYSDDGSISS 618  
QY 189 DREIAKVEQQLILKKKOQOLEEAAKPPPEKPVSPVPEQKHRSIVQIYIDENRKA 248  
Db 619 QSEITGVVQO-----EVEQLDSAGVTGPHPP---SPSEHKVGSSMRPLLOAOLAPSE 669  
QY 249 EAHKIFEGGLPKVPLYNOPSDTKVYHENIKTNQVMRKILILFKRRNHARKQREOKIC 308  
Db 670 RSKPLSSGTGEEQK---QRPHLSLITPEAFDSDDEELEDDEDSAEWRREQQDTA 725  
QY 309 Q-----RYDOLMAWEKKVDRIENNRKAKESKREYIEKQPEIKQREQDERF 359  
Db 726 ESSDDFGSQLRHDYVEDSSEGLSPLPPQPPARAALTTDEDFMRRQILENSAEADNLEE- 784  
QY 360 QRVCQRGAGLSATIA-----RSEHEISELIDGLSEQENNEKQMSQLSVIPPMFMDAQ-- 412  
Db 785 DDTATSGRLAKHGTQGGPRPRPEQPEAPALPKRLPHNATTGYEELLPEGGSAGATD 844  
QY 413 -----RVKFINMN-----GLMEDPMKYKDRQPMVNTDHEKIFDKF 452  
Db 845 GSGTLQGLRRFKTIELNSTGSGYGHLDLQGGDPDS-----LDREPELEMESL 892  
QY 453 IOHPKNFGLTASYLERKSVDPDCLVYLYLTKNENKALVRNRYKGRGRNQOIARPSQEE 512  
Db 893 TGSPED-----RSRGEHSTLTPASTPSYTSSTPSLSEEDSDSPSRQRLEAKQOR 948  
QY 513 KVEEKEDKAEKT--EKKEEKKDEEKKDEKSKENTKEKDKIDGTAETEEREQATPR 570  
Db 949 KARHRSGLPLPTIEDSSEELREELREKQEKMEVEQOIRSTARTR-----R 1001  
QY 571 GRKTANSQGRKGRITRSMTNEAAAAAATAPEPPPLPPPEPPEPPEPPEPPEPPEPPEPPE 630  
Db 1002 DKEELRAQRRE-----RSKT-----PPSNLS--PIEDASPT 1032  
QY 631 EEMEVAKGLVEHGRNWAATAKVGTKSEAKCNFYENYKRRNLDNLLQHQKTKSRKP 690  
Db 1033 ELRQAAE----- 1039  
QY 691 REERDVSCQESVASTVSAQEDIEDIASNEENPEDSEVAVKPSSEDSPEENATSRGNTPEA 750  
Db 1040 MEELHRSSCSEYSPSPSL-----DSAEALDGGPSRLYKSGSEYNLPTF 1083  
QY 751 VELEPTTET--ABSTSPSLAVPTKPAED-----ESVETQVNDISIAE----- 791  
Db 1084 MSLSYPTPTPSGSSSTTPSSGRP--LKSAEAYEEMMRKAELLQROQQAAGARGPHGSPSQ 1142  
QY 792 -----TAEQMDVQOQESHAEEGSCVCDPPPATKADSDVVEVVRVPHENHASKVEDNTKE 843  
Db 1143 PTGPRGLGSEFYQDTTDREYGOAQAAPAEGTASLGAAYEEIL--QTSOSIVMRQASS 1200





QY 1458 -VLRSTLHE-----APKAQLSPGI-YDTSARTRP-----VSYQNTMSRSGSPMMNRRTSDVTI 1507  
DB 2014 MDLSLKHSYSLGFGADGRYLGLOYSVTDLRHPTDLIAHLPLPMRYSVSSVNIYSDHRY 2073  
QY 1508 PPNKSTNHERKSTL-----TPTQRESIPAKSPVPGVDVPSVSHSPFDPHHRGSHGAGEVYWSH 1563  
DB 2074 GPRGDAVGFQENSLAQYSAITAREISRCALNSMD-----OYGRHSGSGGGPDVVOY 2127  
QY 1564 LPTQLDPAMPFHRALDPAAYLFORQL-----SPTGYPSPQYQLYAMENT--RQTILNDY- 1617  
DB 2128 QP-QHGPCGLSAPQSLVPLRGLGNPTPEGHPSFGNLAQYGAAGQTAVRQLLPSTAT 2186  
QY 1618 -----ITSQOMQVNLRPDVARG----- 1634  
DB 2187 VRAADGMIYTIINTPIAATLPTTTPASV-LRPMVRGMYRPAYSAGGITAVPLTSLTRVP 2245  
QY 1635 -LSPREQPLG-----LPYPATRGIIIDLNNPP-----TILVPHPGGTSTPPMD- 1676  
DB 2246 MIAPR-VFLGPTGLYRYPAPSRFTIASSVPPAEGPVYLGKPAAKAPGAGGSPRPEMPVG 2304  
QY 1677 -RITYIFGT-----QITFPPRPYNASMSGPHTHLA-----AAASA 1712  
DB 2305 AAREEPLPTTPAAIKEAAGAPAPLAGQKP--PADAAPGGSGSALSRPGFEKEEASQE 2362  
QY 1713 ERER-----ERER-----EKERERERIA-----AASDLYLRP----- 1740  
DB 2363 ERQRQOQLQLERLEVELEKRLRLQLEELERELRVEQLRHEEQLLVQRELQLOTI 2422  
QY 1741 -----GSEQPRGPGSHGVYRSPSPSVRQETMLQORPSVFGTNGTSTVITPLDPTAQLRI 1795  
DB 2423 KHVIVQQOQEEQAQFALORQLOLEIOLOQLOQL-----QOLEEQQRQK 2474  
QY 1796 MPLPAGGPSISQGLPASRYNTAADA-----LAALVDAASAPO----- 1833  
DB 2475 APFPAACEAPGPGPLAAELAQNGQYWPPLTHAFIAMAGPEGGLQPREPVHLRGLPSS 2534  
QY 1834 -MDVSKTKESHEAARLEENLRSAAYSEQQLBQKT-----LEVERKRSVQC 1880  
DB 2535 ASDMSLQTEQWEASRSGIKKRHSRPLRDACELESGETPCVRRRIADSSVQTDDEGES 2594  
QY 1881 LYTSSAFTSGRPQPHSSVYVSEAGKDKGPPPKSRVEELRTRGKTTIYANFIDVITRQ 1940  
DB 2595 RYLLSRRRRARRSADCVSQTD-----EDSAEQEQVRRRRSR-----LPRH 2636  
QY 1941 IASDKDARERSQSDSSS-----SLSHRYETPSDAEIVSPA-----SSPAPP 1985  
DB 2637 SDGSDSKHDATASSSSAAATVRAMSSVGIQITISCVQTEPDQLPRVSPAIHITAADP 2696  
QY 1986 QEKLOTYQPEVVKANQAE-----DPTROYEG-----PLHHRP----- 2019  
DB 2697 KVEIVRYISAPEKTGRGESLACQTEPDQAQVAGPQLVGPATISPLPGIQTVPGLG 2756  
QY 2020 -QOESPS-----QOOLPPSSQAEGMG-QVPRHRLI-----TLADHICQIITQDFARN 2066  
DB 2757 RFEKKKPDPLEIGYQAHLPPELSQLVSRQPKSPQVLYSPVSPLSPH--RLDTSFSS 2814  
QY 2067 QVSSQTPQOP-----PTSTFQNSPVALVTPVTKTSNRYSPESQAQS 2109  
DB 2815 ERLNKAHVSPOKHFTADSALRQOITLPRPKTLQRS-----LSDPPLSPTAESAKERSL 2870  
QY 2110 VHHQRP-GSRVS-----PENLVKSRGSRGKSPERSHVSSEPIEPIPPQVPVHKEQDSL 2165  
DB 2871 YOHQGLGSOVSALPPNSLVKRVKRTLSPPPPEAHL-----PLAQASQOLVAAS--- 2921  
QY 2166 LLLSQRG-----APPAQRNDARSPGISYLSPSFTKLENTSPMWKSKQEIFRKLNSG 2220  
DB 2922 --LQRLGTGTTVPATKASLLRE-----LDRLRLVEHESTKLKRRQAE----- 2965  
QY 2221 GGDSDMAAQPCTEINPLPAVTTSGSVS-SRGSFADPASNGLIEDIIRKALMGSDFDKV 2279  
DB 2966 --DEEKEIDAKLYLELIGITQRKESIAKDRGGDYPPRLGLGEH---RDYLSSELNQL 3020  
QY 2280 EDHGVYMSQPMGV---VPGTANTSVVTSGETRREE---GDPSPH-----SGGVCK---PKL 2326

DB 3021 RLQG--CTTPAQQFVDFATAAAPATPSGPTAFQOPRFOQPAPQYSAGSGGPTQNGFFPAH 3078  
QY 2327 ISKSNRSRKSPI-----PCOGYLGTERSSVSVSEHSEGDYHRQTPCWANEDRPSSTGSGTQ 2382  
DB 3079 QAPTYGPGSTYPAPAPPAGCASYPABEGLPNOQAFAFRPTGHIYAGOTP-----MPTTOSTL 3131  
QY 2383 FPNY-----PLTMRMLSSPTPTPIACAPSASVNOAAP-----HQ 2415  
DB 3132 FVPADSRAPLQKPRQTSIADLEQKVPINYEVIAS-PVVPMSASAPSETSYSGPVSSGYE 3190  
QY 2416 QNRWE-----REPAP 2426  
DB 3191 QGKVPVPRAGDRGSVSQSPAP 3212

RESULT 41  
Q9N3Y8  
ID Q9N3Y8 PRELIMINARY; PRT: 2344 AA.  
AC Q9N3Y8;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE Y4OC5A.3 PROTEIN.  
GN Y4OC5A.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID-6239;  
RN [1]  
RC SEQUENCE FROM N.A.  
RP STRAIN-BRISTOL N2;  
RX MEDLINE-99069613; PubMed-9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RC SEQUENCE FROM N.A.  
RP STRAIN-BRISTOL N2;  
RA Kalicki J., Randall J.;  
RT "The sequence of C. elegans cosmid Y4OC5A.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RC SEQUENCE FROM N.A.  
RP STRAIN-BRISTOL N2;  
RA Waterston R.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC024772; AAF60538.1;  
DR InterPro; IPR000194;  
DR InterPro; IPR002965;  
DR PRINTS; PR01217; PRICHEXTENS.  
DR PROSITE; PS00152; ATPASE\_ALPHA\_BETA; UNKNOWN\_1.  
SQ SEQUENCE 2344 AA; 7D8696D48D0ADDDB CRC64;

Query Match 3.4%; Score 424; DB 5; Length 2344;  
Best Local Similarity 19.8%; Pred. No. 2.2e-14;  
Matches 457; Conservative 241; Mismatches 807; Indels 806; Gaps 109;  
QY 555 DGTAEETEE-----REQATPRGRKTANSQGRKRGRITRSMTNEAAAAAATAEP- 607  
DB 173 DGDSTSTAEIVLVTTEITTTTSETTTTSAITTEETTEPSTTE--PSTTEETTEET 229  
QY 608 -----PPPLP-----PPPEPISTEPVTSRWTE-- 630  
DB 230 TVSTNPPTPELSDPTDGNATDIDDKIVKKEVSSYTAETPVAPPFFENTVTSVEEPPD 289  
QY 631 ---EEMEVAKGLVEHGRNNAATAKWKYGTSEACKNFYFKRRHNLNLLQHQK-- 684  
DB 290 DDAFEDSKPVKKSTLPNG-----INSKRST-----SYANPLTVDNAAESRWTK 333  
QY 685 -----KTSRKPREE--DVSOCEVASTVSAQEDIEDASNEENPEDEVEAVKP 733







Db 2050 ----PDVKI-----TQHTAEIEDEYLTDAIEIIEHETS--MLTYSEVS--- 2092  
Qy 1230 EGTSPRAHEISLKRKSYESVBNKQKMSMRSPVSPAPLEGLICRALPRGSPHDLAKR 1289  
Db 2093 EGAASILSDASLSSSTSSV-----CTDSSSPIDSGATTGY-----DTSDAVSK 2138  
Qy 1290 TVLGSIMOGTFRATTEFEDGLKYPKQIKRSPPIRAFEAGAITKGPYDGTITTIKEMGR 1349  
Db 2139 LVDSEDIIAQVFTTEEYSE-----VSMYEVSAAGATTKPATASDMDTVHOAAV 2188  
Qy 1350 SIHEIPRODILTOESKRKTEPVVQSTRIEISISOGTPIKFDNNSGQSAIKHNKSLI-T 1408  
Db 2189 CLPE-----TAPSVFTTT--VIKPKQYASDITVIDTAERKDAARKMKSTVET 2234  
Qy 1409 GPSKL-----SRGMPLEIVPENIKVVERGKYEDVKAGET----- 1443  
Db 2235 GIUKTHEDSHKELSDWTRINLTGATSEQPLCLVASVSK-----EPASETPAVP 2285  
Qy 1444 ----VRSRHTSVVSSGP-----SVLRSTLHEAPKAQ-----LSPGI 1475  
Db 2286 TPRVYKSTVSMPSAPALTSKVFSLFRSSLDSP-AQSPSPSPPPPPPPPPPLPPI 2344  
Qy 1476 YDDTSARTPVSYQNTMR-GSPMNRSDVTIPPNKTNHER-----KSTLPTQRESIPA 1531  
Db 2345 LP-----KPAIYPKKKSQIOAPMATAPVPLVTSVATLESAAVLKNHVVPVTKTYTPT 2398  
Qy 1532 KSPVPGVDPVYSHSPDHRGRTAGEVYVWHLPTQL-DPAMPFHRALDPAAAYLFFORQ 1590  
Db 2399 PPPVP---PKSSIP-----AGLVF-SHRPPEVTKPIAPKPAVPPPLPIAVHKPAE 2445  
Qy 1591 LSPTGYPSTQYLYAMENTROTILNDYITSOQMV-----NLRPDVARGLSPREOPLGLP 1645  
Db 2446 TOPKP-----IGLSLTSMTLN-LVSSAEYKIASPISLPHSNKS-SPR---LTKP 2492  
Qy 1646 YPATRGIIDLNMPP--PILVPHPGGTGP-----PM----- 1675  
Db 2493 SOETVWITLSEPGTPTTEAITSOAVTSWPLEAPSKEIQIOPMOPIFTSSMKAVEIQSMA 2552  
Qy 1676 DRITYIPG-----TOITPPRPYNASMSGPHPTHLAAAASAEEREE----- 1717  
Db 2553 DOSMITGALQITPITTOSTTEKVP---SSKSEAVTTEVAKTTASVVRKVPVSPVGLSVT 2609  
Qy 1718 ----REREXERER-----IAAASSDLYLRL-P 1740  
Db 2610 ITIPPEYISQPYRENGRPHGLGDVLDLTLTKVDIEMRSCMDLSAYSMDDARROMP 2669  
Qy 1741 GSEQGRG-----SHGYVRSPSPSVTQETMLQ---QRSPVQGN-----GTS 1782  
Db 2670 TSDTSGRPVSTVQPAIINLSTACVADPSLSIVTETVAVMTCTATVSYGASDLSLDLGH 2729  
Qy 1783 VITPLDPTAQLRIMPLPAGGPSISQGLPASR-----YNTPAADALAVDAASAP----- 1832  
Db 2730 MTTPLQLTTSKHF--PAYRTVSSQPPVSRDEVPINLSLTSAAVTVAAATKPVTPVPV 2787  
Qy 1833 ----QMDVSKTKESKHEAARLENLRSAVAAEQOOLEQKILEV 1873  
Db 2788 SVTNGWTDLSSTQEPMEIGAVDLSTTK-SHRTVVTMDTETSGIITVIEDDE-KPVDLTA 2845  
Qy 1874 EKRSVQC--LY-----TSSAPFGKQPQPH-----SSVYVSEAG-----KDKGPPPKSR 1914  
Db 2846 GRRVACCDVMVYKLPFRGCTAQPPPTLIPDRFGYRDHYQYDRSGSYGYRGMGMKPSM 2905  
Qy 1915 YEEELTRGKTTIAANIDVITRQIASDKDARERGSSDSSSLSSSHRYETPDAIE 1974  
Db 2906 SDTNLSEAGLFAKYSKNSFDYQV-----GATDAAVDLTSGRV-TSGEVD 2949  
Qy 1975 VISPASSAPPOEKQLQTVQPEVWKAQNAENDPTQYEGPLHHYRQBPSPPOQLPSS 2034  
Db 2950 YSSKTTGYP-----ETRQVISGIGISTPQY---SQARWVSSL 2985  
Qy 2035 QAEGMGQVPRHRLITLADHICQIITQDFARNQVSSQTPQOPPTSTQNSPALSVPVR 2094  
Db 2986 SPFGAGSVLRSSNGV-----YSSVATPIPTFAITQPGSIFSTVR 3028

Qy 2095 ----TKTSNRYSPESQA--OSVHHORPGS-----RVS 2120  
Db 3029 DLPTLOTIDSVPSLSTQONQPLPRSYSLFTMAEKDASTLTDIETGLPPLTLESIAE 3088  
Qy 2121 PENLVKSRGRSPKSPERSHVSSEPEYPIPPQVPVY---HEKQDSL-----L 2166  
Db 3089 PTNLIPAT-----ASEVYTDVIEDEVALIIAPECKQOQLDLERELLELEK 3134  
Qy 2167 LLSQGAAPAE-QRNDARSPGISYLPSPFTFKLENSTNTPMWKSKOEI----- 2212  
Db 3135 IKQORFAEELEWERQEIQR-----FREOEKF--MVQKLEELQSMKHLLFQOEE 3182  
Qy 2213 ----FRKL-----NSSGG----- 2222  
Db 3183 ERQAYMMRQETLACQQLQLEQFOOQLOOLHOOLEOEOKIROIYGYDPSGTGSPQTTTD 3242  
Qy 2223 ----DSMAAQAQGTPEINLPAVTTSGS-----VSSRG-----HSFAD 2256  
Db 3243 QALLEGQYATAENG-QFWPTDDATTASGLVGLIEISQSQTWYVQSDGITOYIPRSGILS 3301  
Qy 2257 PASNLGLEDI-----IRK-----ALMGSEDD-----KVEDHGVMVMSQPMG 2291  
Db 3302 SVSEMSLKIDIVREBKQLKRSSMPKLRGYEELEESLEEEPRCYKKIVDSGVQTDDEG 3361  
Qy 2292 VVPGTAN-----TSVVTSGETREEDGDPSPHSGVCKPKL--ISKNSRKS- 2337  
Db 3362 ADRGVTNRRRTKKSVDTSVQTDDED-ODEWDLSSRGR--KPRVGKSESTTEADKAKQ 3418  
Qy 2338 ----PI-----PGQYLG----- 2346  
Db 3419 FSKYSSIAQVVAEISVQTEPVGIRTPSIRARLDAKVEIIKHISAPEKTYKGESLGCQT 3478  
Qy 2347 ----TERPSVSVSHSEGDYHROT---GWAWEDRPSSTGSTOFFYPPNPLMRLMSSTPP 2398  
Db 3479 ETESDTQSQVLSASSQOKDKRPTPLGIGYSSHLRDPDSTLQV-VPSPKPKVLYS-PI 3536  
Qy 2399 TPIACAPSANVQAA--PHQONRIWEREPAPLL 2428  
Db 3537 SPV--SPSKVIESAFVPEYKSTIDDISPOKML 3566

RESULT 43  
Q9H4R4 PRELIMINARY; PRT; 102 AA.  
AC Q9H4R4; 2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE BB329D4.2 (NOVEL PROTEIN SIMILAR TO A TRUNCATED NUCLEAR RECEPTOR CO-REPRESSOR 1 (NCOR1) (RETINOID X RECEPTOR INTERACTING PROTEIN 13)).  
GN BB329D4.2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Babbage A.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL391119; CAC14275.1; -  
KW Receptor.  
SQ SEQUENCE 102 AA; 11336 MW; 462EC7CAAE382014 CRC64;

Query Match 3.3%; Score 414.5; DB 4; Length 102;  
Best Local Similarity 43.7%; Pred. No. 1e-15;  
Matches 90; Conservative 3; Mismatches 4; Indels 109; Gaps 1;

Qy 1 MSSSGYPPNQGAFSTEQSRYPHPHSVOYTFPNTRHQOEFAVPPDYRSHLEYSQASQLLOQQ 60  
Db 1 MSSSGYPPNQGAFSTEQSHYPHPHSKYTFPSTHQQ----- 36

```

Qy 61 OQQQLRRRPSLLSEPHGSDRPOERTSYEPFHGSPVDHDSLESKRPRLEQVSDSHQ 120
Db 37 ----- 36

Qy 121 RVSAVPLVHPLPEGLRASADAKDPAGFKHEAPSPISQPCGDDQNASPSKLSKEE 180
Db 37 -----DPAFGKHEAPSPILGQPCGDDQNASPSKLSKEE 71

Qy 181 LIQSDMRVDREIAKVEOQILKKKKQ 206
Db 72 LIECDMRVDREIAKVEOQILKKKKQ 97

RESULT 44
P70039
ID P70039 PRELIMINARY; PRG; 2829 AA.
AC P70039; P79934;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE ADENOMATOUS POLYPOSIS COLI.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Vlemnickx K., Wong E., Guger K., Gumbiner B.M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U64442; AAB41671.1; -.
DR HSSP; Q02248; 3BCT.
DR InterPro; IPR000225; -.
DR Pfam; PF00514; Armadillo_seg; 6.
DR SMART; SM00185; ARM; 1.
SQ SEQUENCE 2829 AA; 310879 MW; 8A2BABDB77068496 CRC64;

Query Match 3.3%; Score 412; DB 13; Length 2829;
Best Local Similarity 18.6%; Pred. No. 1.3e-13;
Matches 480; Conservative 330; Mismatches 880; Indels 890; Gaps 116;

Qy 169 QNASPSKLSKEELQSDMRVDREIAKVEQILKKKKQOOLEEA----- 213
Db 759 OKALEAEADQHLSETFDNIDNLPKTHR-NKORHKQNLCEYALDSSRHDDSIKSDN 817
Qy 214 -----AKPEPEKVPSPVPEQKHRSIVQIYDENRKAEEAHI 253
Db 818 FSIQNLTVLSPYINTVLPGSSPRPTMDGRPEKDRERTAGLGNVHTTSSGNSKRI 877
Qy 254 FEGLPKVEL-PLYNQPSDTKVYHENIKTNQVMRKKLLILFKRRNHARKQREKICQRYD 312
Db 878 GIQLSTTAQISKVNDVSNHILVQENSSGSA-SEMHCMDSERNRQKPS----- 927
Qy 313 QLMANEKKVDRIENNRKAK-ESKTR-----EYKEQ- 345
Db 928 -----NHQSNPTFTKAESSTRGCPVAFMKMEYKMAINDSLNSVSTEGYKRG 977
Qy 346 --FPEIRKQEQE-RFORVQORGAGLSATARSEH-----EISII 384
Db 978 QVKPSVESYSEDDSKFFSYGYAGLAHKIOSANHHMDNDTELDTPINYSLKYSDEQLN 1037
Qy 385 DGLSEQNEKQOMROLSVPPMDFDAEQRRVKFINMGLMEDPMKVKYKQDFMNVWTDHE 444
Db 1038 SGROSPTONERWRPKHIIDSEMKQEQORP-----TTKTYSSYIENK 1082
Qy 445 KEIFKDFQIHPKNFGLIASYLERKSPDCVLYLYTLTKKNENYKALVRNRYGKRRNQ 504
Db 1083 EEKHK-KF--PPHF-----NQSENVPA-----YTRSGANNO 1111
Qy 505 I-----ARPSQEK-----VEKEEDKAEKTEKKEEKDEKDEKDEKENTKEK 551
Db 1112 VDQSRVSNLNSKASKPHCOVDYDDDKTTF-----SERYSEEQEQEDETQRN---K 1164

```

```

Qy 552 DKIDGTAETEBEQATPRGRK-----TANSOGRKRGTRITSMTEAAAA 596
Db 1165 YNIKAYASEEHGGEOPIDYSRKYSTDVPSSAQKPSFPYNNSSKOKPKKEQVSSNSNPT 1224
Qy 597 SAAAA-----AATEP-----PPPLPPPPPEISTEPVTSRWTEEMEYA-KKGLVE 642
Db 1225 PSPNSNRQNLHPNSAQSRPGLNRPKQIPNKPSPINQETIQT--YCVEDTPICFSRG--- 1279
Qy 643 HGRWAAATAKWGTGKSEAQCKNFYNYRRHNLNLLQOHKOKTSRKPREERDVSQCESV 702
Db 1280 -----SSLSLSSAEDETEGRE--RNSRGOESNNTLQ-----ITEPKISAVSKDGAV 1325
Qy 703 ASTVSA-----QEDIEDIASNEENPEDEVAVKPSDESPENAT--SRGNTEPAVELEP 755
Db 1326 NETRSSVHTTRTKNNRLQTSN-----ISPDSSRHKSVEFFSGAKSPKSGAQ 1373
Qy 756 TTETAP-----STPSLAVPSTKPAEDEVET-QVNDISISAEFAEQMDVDQOBS-----AE 806
Db 1374 TPKSPPEHYVQETPLMFSCRCTSGSLDSFESHSIASSVSA-----SEHMTSGIIS 1426
Qy 807 EGSVCDPPPATKADSDVDVEVRVPENHASKVGEDNTERDLDRASEKVEPRDEDLVVAQOI 866
Db 1427 PSDLPDSPGQTPMPPSRSKTPPPQTVQAKDG-----SKPIVPDEERGKVAK-- 1473
Qy 867 NAQRPEQSDNDSSATCSADEVDGCEPERQRM--FPMDSKPSLLNPTGSIILVSSPLKPNP 924
Db 1474 -----TAVHSAIQRVQVLEADILLHATSTPD-----GFSCASSLSALS 1514
Qy 925 LDLPOLQH--RAAVIPPMVSTPCNIPITGTPVSGYALYORHIKAMHESALL-----BEQ 976
Db 1515 LDEPYIQDVQVLKIMPPVLENDQGN-----KABPEKEFIDNKAKEKDK 1557
Qy 977 RORQEQ-----IDL--ECRSSTSCGTSKSPNRWEVLQAPAPHLITNLPQGVRLP 1025
Db 1558 RSEQEKMDLDDTDDIDILECIISAMP-----RKPSR-----KNKKVP 1596
Qy 1026 TTRPTRPPPL-----IPSKTTVASEK-PSFTMGSGIS-----QGTPTGYTIS 1068
Db 1597 QPTGCKPPPVARKPSQLPVYKLLSSQNRLQTKHVNTSHDDMPRVYCVETGPINFSTA 1656
Qy 1069 HNAQSYTOETP-----KPSVGSISLGLPROQESAKSATLPYIKOEESFSPKSNQSPBG 1121
Db 1657 TSLSDLTIESPPSEPTNDQNTDLSLTLEKRD-----TIP----- 1692
Qy 1122 LTVRAQHEGVVGTAGATQEGSITRGTPTSKISVESIPSLRGSITQGTAPALPQTGIPTEA 1181
Db 1693 -----TEGRSTDODTDASK-----PLNPTTVLDEDK 1717
Qy 1182 LVKGSISRMPEDSDSPKREEAASKGHVYBKGSHILSYDNINKNAREGTRSPRTAHEI 1241
Db 1718 AEEGDILAECIHSAMPK-----CKSHKPYRVKK--IMDQINHHTSAATSSGNSRSMQET 1768
Qy 1242 SLKRSYEV-----EGNIKQGMRESVPVAPLEGLICRALPRGSPHS----- 1284
Db 1769 DNKKTSPVKPMQPSIGFKERLKNTELKNPNS--ENQYCD--PR-KPSSKKPKSKVAN 1822
Qy 1285 ---DLKERTVLSG-----STMOCTPRATTES-----FEDGLKYPKQIKRESPP 1324
Db 1823 EKIPNNEERT--KGFAFDSPHHYPTIEGTPYCFSRNDSLSLDFEDD---DIDLSEKAE 1877
Qy 1325 IRAFEGAITKGPYDGIITIKEMGRSIIHEIPRODILLTOES---RKTPE-VVO----- 1372
Db 1878 LRKEKGT---KDTQKVKYKHENRAINPMGKQDQTPKSGLGRRQOPKALQKPTSFSSA 1933
Qy 1373 -----STRPIEGSIGQTPIKFDNNSSQSAI-----KHNYKSLITGFSKLSRGM 1418
Db 1934 AKGTQDRGADKDEKMFENAIENTPVCFSRNSSLSDIDQENNKKE--TEPLKQT---- 1987
Qy 1419 PLEIVPENIKVVERGKYEDVKAETVRSRH---TSVSSGGFSLSTLHEAPKAOLSPGI 1475
Db 1988 -----GTSETOLGLRRPQTSYGAPKPSFHVEDTPVCFSRNSSLSDSDSDLLQECI 2040

```

QY 1476 YDDTSARRTPVSYQNTMSRG-----SPMNRTSDVTIPPNKSTNHERKSTLTPTQRESIP 1530  
Db 2041 SSAMPKKRPSKIKVEGKSRNSVGGIIAEEPDITLDRIOQSDSENAFSP-DSENF 2099  
QY 1531 AKSPVGVDPVSHSPFPHHGRSGAGEYVWS-----HLPTQLDPAMPFHRALD-- 1579  
Db 2100 WKAIQEGANSIVSR-----HQAAGSLSRQSGSDSILSKSGISLGPFLHLDKE 2154  
QY 1580 -----PAAAYLFORQLSTP-----GYPOYQLYAMENTROITLNDYIT-- 1619  
Db 2155 EKTITSNGPKILKPAKSALENNKTEEPKIGKGVKYSKITGKRS--SDFSHC 2212  
QY 1620 SOOMVNLRPDVARG-----LSPREOPLGLPYPATRGIIIDLTNMPPTILVPHG 1668  
Db 2213 KQSVQTNM-PSISRGRTMIHHPGVASSPSTSPVSKKGPFVK-----NVPSKSGNEPS 2265  
QY 1669 GTSPTPMDRITVPGTQITFPPRPNYSASMGPHTHLAAASABREREREKERER 1728  
Db 2266 SSSSP-----KGT-----KPLKSELVYGRSPSTPGGSKGNSRSGSRD----- 2304  
QY 1729 IAAASDLYLRPGSEOPGRPGSHGVRSFVSRTQETMLQORPSVFOGTNGTIVITLD 1788  
Db 2305 --SASS-----RSPQPLSRP-----LQSPG-----RNSISPGKNG--ISPPN 2338  
QY 1789 PTAQLRIMPLPAGGPSISQGLPASRYNTAADAALAAVDAASAPQMDVSKTESKHEAAR 1848  
Db 2339 KFSQLPRTSPSTASTKSG----- 2358  
QY 1849 LEENLRSAVSEGOQLEQKLEVEKRSVQCLYTSSAPPGSKQPHSHSVVYSEAGDKG 1908  
Db 2359 -----SGRMSYTSRQOLSQPNLSKQ-----SGLPKTHSSIPRESAS-- 2396  
QY 1909 PPKRYEEELTKTKTITTAANFIDVITROIASDKDARERGSSDSSSLSSHRYET 1968  
Db 2397 -----KSLNQNTGNS-----KKV-----ELSRMSYTSKSGESDRSER 2431  
QY 1969 PSDAIEVISPASSAPPOEKLQYQYEVVYKANAENDPTROYEGHLHYRPOQESPSP-- 2026  
Db 2432 PA-----LVROSTFK----- 2449  
QY 2027 QOOLPPSSQAEGMGVPRTHRLITLADHLCIITODFARNQVSSOTPOOPTSTFQNSPS 2086  
Db 2450 RKLEESASFELS-----SSRADSPRPSQTQ-TPA 2480  
QY 2087 ALVSTPVRTKTSNRYSPESQAQSVHH-ORPGSRVSPENLVDSKRSRPGKSPERHVSSE 2145  
Db 2481 LSPSLP-----DMALSTHSIQAGGWKPPNL---NPAEHGDSRRHDISRS 2525  
QY 2146 PYEPTSPQVPPV-----HEKQDSLLLSQRGAEPAEQNDARSFGSISYLPSTFTK 2197  
Db 2526 HSE--SPGRLPITRSGTWKREHSHSSSL-----PRVSTWRTTSGSSSILS---- 2569  
QY 2198 LENTSPMKVSKQEIFKRLNSSGGSDMAAQAQCT-----ELFNLPAVTTSGSVSR 2250  
Db 2570 -ASSESSKASEDEKQVCSPGPRSECCSAGKTWKRIESELTP-----SN 2619  
QY 2251 GHSFADPASNLGLEDIIRKALMGSDDDKVEDHGV-----VMSQPMGVVPGTANTSVWTS 2305  
Db 2620 GSSSTIAESNCLESKTLVQYMAPAVSKTEDVWVRIEDCPINNPSRSGSPICNSPPVIDN 2679  
QY 2306 -----ETFRREGDSPH-----SGGVCKPLKLSKNSRKSIPGQGYLGTERTSPSSVSHS 2358  
Db 2680 VLDQCKEAAKDCHTRNSNGNVPFL---ENROKSKIKVDGLDTKTGD-PKSLINQQ 2735  
QY 2359 EGDYHQTPGAWEDRPSGTGS-----TQPPYPLTMRM-----LSSTP---PPI 2401  
Db 2736 ETNENTVAERTAFSSSSSKSHSPSGTVAARTVPPNYNAPRKSNGENSTSRPSQIPTPV 2795

RESULT 45  
Q94226  
ID Q94226  
AC Q94226;

PRELIMINARY; PRT; 2361 AA.

DT 01-FEB-1997 (TREMblrel. 02, Created)  
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)  
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
DE SIMILARITY TO SALMO GAIADNER.  
GN F45E4.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID-6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Wilson R.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Waterston R.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U70852; AAB09135.1; -;  
SQ SEQUENCE 2361 AA; 261579 MW; 75F7106658D38E3B CRC64;

Query Match 3.2%; Score 409.5; DB 5; Length 2361;

Best Local Similarity 18.6%; Pred. No. 1.4e-13;

Matches 523; Conservative 374; Mismatches 975; Indels 943; Gaps 138;

QY 79 SDRPOERTSYEPHPGPSVDHGSLEKPRLEQVSD--SHFORVSAVLPLVHPLPEG 136  
Db 22 SDRPSALAMF---TNPSASPSL--LRKESEDRGDIILNLRSSGA-----D 65  
QY 137 LRASADAK---KDPAGFGKHEAPSPISGQPCGDDQNASP-----SKLSKEBELIQSM 185  
Db 66 SRASNDSSASRLPDTALVGLSEAEKQHTMSVMSRSNRNTSPMTSRRCSSALQMLPEVDNL 125  
QY 186 DRVRE---IAKVEQ---ILKLKKQ---QLEEEAAKPPPEKP 222  
Db 126 SEAEKEHTQTLKAEKSTPFMIKVPKMKQISSRTESNSRVSSGEIDEEVENEVQRKT 185  
QY 223 VSPPPVEQKHSIVQIYDENKKAEEAHKIFEGLPKVEL---PLYNQPSDTKVYHEN 278  
Db 186 IEEPIEIPSAVT---PRNLRV-----IPPIAISHPTTPPSAKTDTGSRHS 232  
QY 279 -----IKTNQVWRKKLILFFKRRNHAKQROKICQRYDQLMEAWKVKVDRIENNP 329  
Db 233 GSSAHSQFGFTSPISGFK--IFFDK--AKTATETLKEIKDEIVIVVDK--DKTETK 285  
QY 330 RKAKESKTRVEYKQFPIRKQEQERFORVQORGAGLSATI---ARSEHEISEID 385  
Db 286 EPNVSNELTAE---EFEHIRINEM-----AGIDEPIQPPPIAERRKSSVVS 330  
QY 386 GL-----SEQENNEKO-MRQLSVIPPMFDAQRRVKFINMGLMDPMKVYKD 433  
Db 331 GLKNMFGVGKHDESELTTTEKEHIRM--LMAEK----- 364  
QY 434 RQFMVWTDHEKEIFKDFKFIQHPKN-FGLIASYLERKSVPCVLYLYLTKNENYKALVR 492

```

Db 365 -----DEEILDEQ--SKPKSTFGL-----KNF-----384
QY 493 RNYGK-----RRGNQOIARPSQEEKVEEKEEDKAETKEEKKDEEEDK 539
Db 385 --FGRATOSVHMATDSVINKVQNO--SKQSLGLTQEBLNDIAQATSAQAQESKQELTQD 440
QY 540 EKEDSKENTKEDKIDGTAETEEREQATPRGRKTPANSQGRKRGRITRSMTNEAAAAAA 599
Db 441 EIDHIA-----RINMAAEDFEQPAVN-----MNQ-----LSQAEKDHIAIE 479
QY 600 AAAATEEPPPLPPPEPISIPETVETSRWTEEMEVAKGLVEHGRNAAIAKAVGTKSE 659
Db 480 AWAADDSKIIIVPPPSIIQKD--IELM--EEMD-----HIARIAAMAD-----520
QY 660 AQCKNFYFNKRRHNLNLOOHKOKTSRKKPREEDVSOCE-----SVASTVSAQED 711
Db 521 ---EDF-----THPVKGAAPYIDENYSRDDGAIDRFPASTATPVFAQPS 562
QY 712 EDIEASNEEENPEDSEVBAVKPSEDPENATSRGNTEPA--VELEPTTE-----TAPS 762
Db 563 E-IELS--BEEREHISRIAMAEEDFN---TPYVSTHTPTQOIEIETEEKEHIAIEAM 616
QY 763 TSPSLAVSTKPAEDSEVETO---VNDISAEATQMDVDQOEHSABEGSVCDPPATKA 819
Db 617 ASEDLNAPSPTFQOQRVATMPVPINYHVEEPTLSQEEID---HIARITAM-----ATE- 667
QY 820 DSDVDVEVRPNHASKVEGDNTERDORASEKVEPRDEDLVVAQOINAAQPEQSDNDS 879
Db 668 DNTDLOT-LPTPOVKQNEPELSQB--EIDHITRIAMANEFGMYSIVSEHPAPV--ES 722
QY 880 SATCSADEEDVGEPRQRMFMDSKPSLLNPTGSLVSPKPNPLDLPQ--LOHRAAVI 937
Db 723 ELT---QEELDHIAKITGMA5MDIS--TLPPPTG-----KPSETSLTQEBLHIARIA 770
QY 938 PPMVSCTPCNIPICPTVSGYALYQRIKAMHESALLERQORQOEIDLECRSSTSPCGTS 997
Db 771 E--MASAEYDVPV-----KIFEPPELTQEBLHIAKITAMAA---805
QY 998 KSPNREWEVLOPAPHOLITNLPEGVRLTPTRPPPLIPSSKTTVASEKPSFTMGSI 1057
Db 806 -----QDVOLPATQRSITH-----KVSITPP--PPPP-----830
QY 1058 SQGTPTGYLTSHNOASYTQETPKPSVGSISIGLPRQOESAKSATLPY-----IKOPE 1109
Db 831 -----SKHFEQDLTQEE-----LOHARIAEMADMDYNTPTTAEVPQOEE 870
QY 1110 FSP--RSONSQOEGLLVRAOHEGVVGRGTAGALQOEGSITRGPTSKISVESIPSLRGSTQ 1167
Db 871 EEPITESGSDATSGADIDFQEDASGSGSFGDNNAQ--VLTSGFSPDRVTSFAALDTT 928
QY 1168 GTPALPQTGIPTEALVKGISGRMPTEDSSPEKGREEAASKGHVIEGKSGHTLSYDNLKN 1227
Db 929 EEPQCP-----TMAQTVSFTPSADSMARKSSE-----YD-----IRSISEIRQ 968
QY 1228 ARECTRSPTAHEIS-LKRSYSEVGNIKOGMSRSPVSAPLEGL-----ICRALPRGS 1281
Db 969 ESESDIGKWEQSLFMSRQSIHDEEDV--GHEINTDVEEFPLEVEEQHLHFLEGIDVES 1026
QY 1282 PHSDLKERTVLSG---STMQ-----TPRATTSEFEDGLKYP 1315
Db 1027 QHHEPTTSSAFGTGRSITGEGERKKYGDGDAVEQOQKLENYEEEEKTKSSREAFDDGFETQ 1086
QY 1316 KO--IKKRESPPIRAFAEGAITGKP-----YDGIITIKEMGRSIIH-EIPRODILTQE 1363
Db 1087 REESLRAQOTFIDSLPGSRMLKRNFGPLSIANDAINKAEAGSQIAQAAVPIRPSSSS 1146
QY 1364 SRKTPEVVQSTRPIEGSISQGT--PIK-----FD 1391
Db 1147 NIVNNVFSKK-----SSTSLGTSAPTAKSIPSPQIGIPMDGLSEERQIOMSVMAAADFD 1202
QY 1392 NNSGOSAIKHNKVLITGPKSLSRGMPPLEIVPENIKVVERKGYDVKA-----GETVR 1445

```

```

Db 1203 DS-----VNNYKPTSGSSNIPAGM-----EDLSEAEEREKIMSVMAEMEMGAREP 1249
QY 1446 SRHTSVVSGPSPVSLRSTLHEAPKAOLSPGIYDDTSARRTPV-----SYQNTMSRG 1495
Db 1250 PPSOIIPTRSPSVMSSSI-----MSELPGGLDLSDDERKIMAVMAEADQNVKPIAG 1305
QY 1496 SPMNRTSDVTIPPN-----KSTNHERKSTLTPTQRE 1527
Db 1306 PPMMPSTSM-IPPGMEGLSEERQKIMSVMAAETDSSSVITSRQPSRSPSVARMQPO 1364
QY 1528 STPAKSPVPGVDP-----VVSHPDPHHRGSTAGEVYWSHLPTQLDPAM 1572
Db 1365 LMPQQAIPPIIPPLEGLSDEERHTIMSVMAEAEFEESR--SQVPSRQSRSPSFVNQO 1422
QY 1573 PPHRALDPAAAAYLFQROLSPTPGY-----PSQOLYAMENTROTILNDYITSQOMOVNL 1627
Db 1423 SFH-----PIPSFPIVPPGLEDL--SEERQKIMS-----VWMA 1456
QY 1628 RPDVARGLSRQPIGLPYPATRGIIDLTMNPPITILVPHPGGTSTPPMDR---ITYIPT 1684
Db 1457 EVEESRQOLPSRQPSKPSVA-----MIQAPAVPIIP-SGMEDLPEAERQKIMSVMAEA 1509
QY 1685 QIT--FPPRPYNSAKSPGH--PTHAAAAAERER--EREREKERERERIAAASDLYL 1738
Db 1510 EQNFVPSRSPSNYSMQPVPIPHGLEDLSEAEERKILSVMAEAEIDSAKIPSRSTSYS 1569
QY 1739 R-----PGSEQP-----GRPGSHGVRSRSPSVRTOETMLQORPSVFOQT 1778
Db 1570 MPPPLPQMSQPELITGLEHSSADMEFGDSSRSRHOVIPPGLLENSEERQKIMSVMAHA 1629
QY 1779 NGTSVITLDPDTAQL-RIMPLPAGGSPISQGLPASRYNTAADALAAALVAAAASAFQMDYS 1837
Db 1630 ELESII---PSGHIDQIPISLPRGHTGFK---PAGIVN--EDEL-----FETERKOREES 1675
QY 1838 KTKESKH-----EAARLEENLSR-----1856
Db 1676 PYRESGYATSYERELAMGDEERMDGLELIIIRREGARSRRDSRDEVLHRRREEDPEVH 1735
QY 1857 -----SAVSEQOOLEKLTLEKRSV---QCLYTSSAFP-----1888
Db 1736 TPEESTAVVIDVPSVSPVTENVPEKQTDFFTFYSDSRFAEIVOMQEEEEAGSLQKQV 1795
QY 1889 SGKPO-----PHSSVYVSEAGKDKGPP---PKSRYE-----EELTRGK 1924
Db 1796 DEKPRMETVFDGSELSLPHODEFVENEPTTKTSDFPDKETDEVFEKPEIQRIVRTKN 1855
QY 1925 TTITAAINDIITIIQIADSKDARERGSSQSSSSSLSH-----RVE 1967
Db 1856 HDVDMDEIYDNVIAEAPSSVSQRR---QPVDSSETSVKSRSTIQRGTPKPPMIKITVEEE 1912
QY 1968 TPSDAIE-----VISPASSAPPOEKLTQYQEVVKKANQANDPTROYEGL 2014
Db 1913 TKSDSESCSEDDDEYDPRVVAAPTAPATFE-----EVENERIQ-----1954
QY 2015 HHYRQOESPSPQQLPPSPQAEGMGQVPRHLITLADHICQIITQDFARNQVSSQTPQ 2074
Db 1955 -----EELGKEVLOQI---MAFG-----EVANDEFDVQWAKTTISQ 1987
QY 2075 QPPTSTFQNSFALSALYSTPVRTKTSNRYSPESQAQSVHHQPGSRVSPENLVDKS---RGS 2131
Db 1988 TPSTSTKPTVTAPKRS DPI-----PIAPSQRKSKEIEE---RIRTEALEEEEFYRHGH 2037
QY 2132 RPK-KSPERSHVSSEPEYIPSPQVPVVEKHODSLLLSQGAEP-----EQNDARS 2184
Db 2038 NPFLESPEDEVSIN-MEDVEYAEIARLYESANQ---TMRPPGVYVITTEDESDDGLT 2092
QY 2185 PGCSILYPSFTKL--ENTSPMVKSKQOEI FR-KLNSGGGSDGDMAAAQPGTEIFENLPAV 2241
Db 2093 SNSESRLVAREKRLMDKKTADSLMAKYQMKKVKQAQKQTTSSTVATPAIINFSDL 2152
QY 2242 TTSQSVSRGHSFADPASNLGLEDI-----IRKALMSGFDDKVEDGHVYMSQ-----2288
Db 2153 KTS-TRTDSNSYFETTKNIPALEIKDPKPIDPPEISASIDKTMAEVDALLQOYVNEKA 2211

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 14:39:33 ; Search time 19.61 Seconds  
(without alignments)  
4262.284 Million cell updates/sec

Title: US-09-522-753-11  
Perfect score: 12643  
Sequence: 1 MSSSGYPNQGAFSTEQSRY.....EREAPPLISAQVETLSDSD 2440

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score   | Query Match % | Length | DB ID | Description |
|------------|---------|---------------|--------|-------|-------------|
| 1          | 12643   | 100.0         | 2440   | 1     | NCRI_HUMAN  |
| 2          | 11336.5 | 91.2          | 2453   | 1     | NCRI_MOUSE  |
| 3          | 4147.5  | 32.8          | 2517   | 1     | NCRI_HUMAN  |
| 4          | 4055.5  | 32.1          | 2472   | 1     | NCRI_MOUSE  |
| 5          | 2493.5  | 19.7          | 533    | 1     | NCRI_RAT    |
| 6          | 480     | 3.8           | 2468   | 1     | MAPB_HUMAN  |
| 7          | 465.5   | 3.7           | 2774   | 1     | MAPA_RAT    |
| 8          | 447     | 3.5           | 2464   | 1     | MAPB_MOUSE  |
| 9          | 437.5   | 3.5           | 2805   | 1     | MAPA_HUMAN  |
| 10         | 437     | 3.5           | 2845   | 1     | APC_MOUSE   |
| 11         | 411.5   | 3.3           | 3924   | 1     | ANK2_HUMAN  |
| 12         | 406     | 3.2           | 2842   | 1     | APC_RAT     |
| 13         | 404     | 3.2           | 3256   | 1     | K167_HUMAN  |
| 14         | 390     | 3.1           | 1226   | 1     | YCS3_YEAST  |
| 15         | 374.5   | 3.0           | 3562   | 1     | PGCV_CHICK  |
| 16         | 365     | 2.9           | 2843   | 1     | APC_HUMAN   |
| 17         | 360     | 2.8           | 2349   | 1     | TPR_HUMAN   |
| 18         | 353     | 2.8           | 2142   | 1     | BAT2_HUMAN  |
| 19         | 350     | 2.8           | 661    | 1     | YDBJ_SCHPO  |
| 20         | 348.5   | 2.8           | 1018   | 1     | YKZ6_CAEEL  |
| 21         | 343.5   | 2.7           | 1411   | 1     | TCOF_HUMAN  |
| 22         | 343     | 2.7           | 3866   | 1     | HRX_MOUSE   |
| 23         | 336     | 2.7           | 2688   | 1     | ZEPL_MOUSE  |
| 24         | 332.5   | 2.6           | 1861   | 1     | MAP2_RAT    |
| 25         | 326     | 2.6           | 1781   | 1     | AKAC_HUMAN  |
| 26         | 321.5   | 2.5           | 3969   | 1     | HRX_HUMAN   |
| 27         | 317     | 2.5           | 4687   | 1     | PLEI_RAT    |
| 28         | 314.5   | 2.5           | 3358   | 1     | PGCV_MOUSE  |
| 29         | 312     | 2.5           | 1902   | 1     | SMEL_HUMAN  |
| 30         | 310.5   | 2.5           | 2004   | 1     | MOZ_HUMAN   |
| 31         | 309.5   | 2.4           | 2441   | 1     | CBP_MOUSE   |
| 32         | 308     | 2.4           | 2442   | 1     | CBP_HUMAN   |
| 33         | 303     | 2.4           | 3396   | 1     | PGCV_HUMAN  |

34 300.5 2.4 2090 1 HFCL\_MESAU  
35 299.5 2.4 2715 1 TRX2\_HUMAN  
36 298.5 2.4 771 1 CALD\_CHICK  
37 298 2.4 2194 1 SC16\_YEAST  
38 296.5 2.3 1828 1 MAP2\_MOUSE  
39 295.5 2.3 848 1 NFM\_MOUSE  
40 295.5 2.3 2035 1 HFCL\_HUMAN  
41 295 2.3 2717 1 ZEPL\_HUMAN  
42 294 2.3 915 1 NFM\_HUMAN  
43 291 2.3 1827 1 MAP2\_HUMAN  
44 289 2.3 1637 1 MRSP\_STAAU  
45 284.5 2.3 1505 1 CDP\_HUMAN

## ALIGNMENTS

RESULT 1  
NCRI\_HUMAN  
ID NCRI\_HUMAN STANDARD; PRT; 2440 AA.  
AC 075376; Q9UPV5; Q9UQ18;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR).  
GN NCOR1 OR KIAA1047.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal brain;  
RX MEDLINE=98393736; PubMed=9724795;  
RA Wang J., Hoshino T., Redner R.L., Kajigaya S., Liu J.M.;  
RT "ETO, fusion partner in t(8;21) acute myeloid leukemia, represses  
RT transcription by interaction with the human N-Cor/msin3/HDAC1  
RT complex";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).  
RN [2]  
RP SEQUENCE OF 782-2440 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=99397452; PubMed=10470851;  
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,  
RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIV.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:197-205(1999).  
RN [3]  
RP SEQUENCE OF 974-2440 FROM N.A.  
RX MEDLINE=99375326; PubMed=10444336;  
RA Nagaya T., Chen K.-S., Fujieda M.,  
RA Horwitz K.B., Lupski J.R., Seo H.;  
RT "Localization of the human nuclear receptor co-repressor (N-COR) gene  
RT between the CMT1A and the SMS critical regions of chromosome  
RT 17p11.2";  
RL Genomics 59:339-341(1999).  
CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES  
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE  
CC ABSENCE OF LIGAND.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
CC DOMAINS (ID1 AND ID2).  
CC -!- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND

SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES  
FLANKING THE CORN BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
SPECIFICITY.  
-!- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A-SANT-B).  
-!- SIMILARITY: CONTAINS 2 CORN BOX.  
-!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF044209; AAC33550.1; -;  
EMBL; AB028970; BAA82999.1; -;  
EMBL; AB019524; BAA75814.1; -;  
MIN; 600849; -;  
InterPro: IPR001005; myb.DNA-binding; 2;  
Pfam: PF00249; myb.DNA-binding; 2;  
PROSITE; PS50090; MYB\_3; 1;  
Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
coiled coil.  
KW COILED COIL.  
FT DOMAIN 174 216 COILED COIL (POTENTIAL).  
FT DOMAIN 254 312 INTERACTION WITH SIN3A/B.  
FT DOMAIN 299 328 COILED COIL (POTENTIAL).  
FT DNA\_BIND 437 482 SANT-A (POTENTIAL).  
FT DNA\_BIND 625 670 SANT-B (POTENTIAL).  
FT DOMAIN 501 557 COILED COIL (POTENTIAL).  
FT DOMAIN 607 617 PRO-RICH.  
FT DOMAIN 988 1816 INTERACTION WITH ETO.  
FT DOMAIN 2055 2059 CORNR BOX OF ID1.  
FT DOMAIN 2263 2267 CORNR BOX OF ID2.  
FT DOMAIN 58 64 POLY-GLN.  
FT DOMAIN 593 603 POLY-ALA.  
FT DOMAIN 1032 1035 POLY-PRO.  
FT DOMAIN 1707 1712 POLY-ALA.  
FT DOMAIN 1952 1963 POLY-SER.  
FT CONFLICT 1014 1014 L -> V (IN REF. 2).  
FT CONFLICT 1508 1509 PP -> SS (IN REF. 2).  
FT CONFLICT 1561 1561 W -> R (IN REF. 2).  
FT CONFLICT 1567 1567 Q -> H (IN REF. 2).  
SQ SEQUENCE 2440 AA; 270263 MW; 60A4D7964D00EDAB CRC64;

Query Match 100.0%; Score 12643; DB 1; Length 2440;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2440; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSSSGYPNQGAFSTEQSRYPHVSQVTFPTNRHQEQFAVPDYRSSHLEVSQAQLQQQ 60  
DB 1 MSSSGYPNQGAFSTEQSRYPHVSQVTFPTNRHQEQFAVPDYRSSHLEVSQAQLQQQ 60  
QY 61 QQQLRRRPSLLSEFFHSGDRPQERRTSYEPFHGSPVDHDSLESKRPRLEQVSDSHFQ 120  
DB 61 QQQLRRRPSLLSEFFHSGDRPQERRTSYEPFHGSPVDHDSLESKRPRLEQVSDSHFQ 120  
QY 121 RVSAVPLVHLPEGLRASADAKDPAGFGKHEAPSSPIGQPCGDDQNAPSLSKEE 180  
DB 121 RVSAVPLVHLPEGLRASADAKDPAGFGKHEAPSSPIGQPCGDDQNAPSLSKEE 180  
QY 181 LIQSDMRDREIAKVEQOILKKKQOOLEEAAKPEPEKVPSPVPVQKHRSIVQI 240  
DB 181 LIQSDMRDREIAKVEQOILKKKQOOLEEAAKPEPEKVPSPVPVQKHRSIVQI 240  
QY 241 DENRKAEEAHKIEFGLGPKVELPLYNQPSDTKYVYHENIKTNQVARKKLIFFKRRNHAR 300  
DB 241 DENRKAEEAHKIEFGLGPKVELPLYNQPSDTKYVYHENIKTNQVARKKLIFFKRRNHAR 300  
QY 301 KORQOKICQRYDQLMEAWKVKVDRIENNRKAKESKTREYIEKQFPEIRKQEQBERFQ 360  
DB 301 KORQOKICQRYDQLMEAWKVKVDRIENNRKAKESKTREYIEKQFPEIRKQEQBERFQ 360

DB 301 KORQOKICQRYDQLMEAWKVKVDRIENNRKAKESKTREYIEKQFPEIRKQEQBERFQ 360  
QY 361 RVGORGAGLSATIARSEHEISEIIDGLSEQENNEKQRLQSVIPPMFADARQVRKFNNM 420  
DB 361 RVGORGAGLSATIARSEHEISEIIDGLSEQENNEKQRLQSVIPPMFADARQVRKFNNM 420  
QY 421 NGLMEDPMKVKYKDRQFMNVWTDHEKEIFKDKFIOHPKNFGLIASYLERKSPDCVLYYL 480  
DB 421 NGLMEDPMKVKYKDRQFMNVWTDHEKEIFKDKFIOHPKNFGLIASYLERKSPDCVLYYL 480  
QY 481 TKKNENYKALVRRNYGRRGRNQIARPSQEEKVEEKEEDKAETKEEKEEKEDEKDE 540  
DB 481 TKKNENYKALVRRNYGRRGRNQIARPSQEEKVEEKEEDKAETKEEKEEKEDEKDE 540  
QY 541 KEDSKENTKEKDKTDGTAETETEEEOATPGCKTANSGRKGRTBSMTNEAASAAA 600  
DB 541 KEDSKENTKEKDKTDGTAETETEEEOATPGCKTANSGRKGRTBSMTNEAASAAA 600  
QY 601 AAATEEPPLPPPPPEPISTEPVETSRWTEEMEVAKKGLVEHGRNNAIAKMYGTKEA 660  
DB 601 AAATEEPPLPPPPPEPISTEPVETSRWTEEMEVAKKGLVEHGRNNAIAKMYGTKEA 660  
QY 661 QCKNFYFNKRRHLDNLQHQKOTSRKPREERDVSOCEVASTVSAQEDEDIASNEE 720  
DB 661 QCKNFYFNKRRHLDNLQHQKOTSRKPREERDVSOCEVASTVSAQEDEDIASNEE 720  
QY 721 ENPEDSEVAVKPSEDSPENATSRGNTPEPAVELEPTTETAPSTSPSLAVPSTKPAEDSV 780  
DB 721 ENPEDSEVAVKPSEDSPENATSRGNTPEPAVELEPTTETAPSTSPSLAVPSTKPAEDSV 780  
QY 781 ETQVNDISIAETAEQMDVDQOEHSAAEGSVCDPPATKADSDVDVEVRVPENHASKVEGDN 840  
DB 781 ETQVNDISIAETAEQMDVDQOEHSAAEGSVCDPPATKADSDVDVEVRVPENHASKVEGDN 840  
QY 841 TKERDLDRASEKVEPREDELVVAQIINAQRPESQDNDSSATCSADEVDGEPERQRMFP 900  
DB 841 TKERDLDRASEKVEPREDELVVAQIINAQRPESQDNDSSATCSADEVDGEPERQRMFP 900  
QY 901 MDSKPSLLNPTGSLVSSPLKPNLDLPOLQHRAAVTPMVSCPTPCNIPICPTVSGVALY 960  
DB 901 MDSKPSLLNPTGSLVSSPLKPNLDLPOLQHRAAVTPMVSCPTPCNIPICPTVSGVALY 960  
QY 961 QRHIKAMHESALLEEQRQEQIDLECRSSTSPCGTSKSPNREWEVLQAPAPHLITNLE 1020  
DB 961 QRHIKAMHESALLEEQRQEQIDLECRSSTSPCGTSKSPNREWEVLQAPAPHLITNLE 1020  
QY 1021 GVRLLPTRPRPPPLIPSSKTTVASEKPSFIMGGSISQGTPTGYLTSHNQASTQETPK 1080  
DB 1021 GVRLLPTRPRPPPLIPSSKTTVASEKPSFIMGGSISQGTPTGYLTSHNQASTQETPK 1080  
QY 1081 PSVGSISLGLPROQESAKSATLPYIKOEFSRSONSQPEGLLVRAOHEGVVRCGTAGAIQ 1140  
DB 1081 PSVGSISLGLPROQESAKSATLPYIKOEFSRSONSQPEGLLVRAOHEGVVRCGTAGAIQ 1140  
QY 1141 EGSITRGTPTSKISVESIPSLRGSITQGTALPOTGIPTEALVKGSIIRMPIDESSPEKG 1200  
DB 1141 EGSITRGTPTSKISVESIPSLRGSITQGTALPOTGIPTEALVKGSIIRMPIDESSPEKG 1200  
QY 1201 REEAASGHVIEYEGKSHILSYDNINKNAREGTRSPRTAHEISLKRSEYVEGNTKQGM 1260  
DB 1201 REEAASGHVIEYEGKSHILSYDNINKNAREGTRSPRTAHEISLKRSEYVEGNTKQGM 1260  
QY 1261 RESPVSAPLLEGLICRALPRGSPHSDLKERTVLSGIMQGTPTATTESFEGCLKYKPKOIKR 1320  
DB 1261 RESPVSAPLLEGLICRALPRGSPHSDLKERTVLSGIMQGTPTATTESFEGCLKYKPKOIKR 1320  
QY 1321 ESPPIRAFEAGITKGPYDGTITTIKEMGRSITHEIPRODILTQESRKTPEVYQSTRPIEG 1380  
DB 1321 ESPPIRAFEAGITKGPYDGTITTIKEMGRSITHEIPRODILTQESRKTPEVYQSTRPIEG 1380  
QY 1381 SISOGTPIKFDNNSGQSAIKHNKSLITGPKSLRGMPPELIVPENIKVVERGKYEDVKA 1440  
DB 1381 SISOGTPIKFDNNSGQSAIKHNKSLITGPKSLRGMPPELIVPENIKVVERGKYEDVKA 1440









DR InterPro; IPR001005; --  
 DR pfam: PF00249; myb-DNA-binding; 2.  
 DR PROSITE; PS50090; MYB\_3; 1.  
 KW Nuclear protein; transcription regulation; DNA-binding; Repressor;  
 KW coiled coil; Alternative splicing.  
 FT DOMAIN 174 215  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 254 312  
 FT INTERACTION WITH SIN3A/B (BY SIMILARITY).  
 FT DNAS\_BIND 429 474  
 FT SANT-A (POTENTIAL).  
 FT DNAS\_BIND 612 657  
 FT SANT-B (POTENTIAL).  
 FT DOMAIN 522 561  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 778 820  
 FT PRO-RICH.  
 FT DOMAIN 2139 2143  
 FT CORNR BOX OF ID1.  
 FT DOMAIN 2342 2346  
 FT CORNR BOX OF ID2.  
 FT DOMAIN 494 510  
 FT POLY-GLN.  
 FT DOMAIN 682 685  
 FT POLY-LYS.  
 FT DOMAIN 994 1002  
 FT POLY-PRO.  
 FT DOMAIN 1384 1389  
 FT POLY-PRO.  
 FT DOMAIN 1842 1846  
 FT POLY-GLY.  
 FT DOMAIN 2479 2482  
 FT POLY-PRO.  
 FT VARSPLIC 1 1702  
 FT MISSING (IN ISOFORM TRAC-1).  
 FT VARSPLIC 2353 2398  
 FT MISSING (IN ISOFORM TRAC-1).  
 FT CONFLICT 7 7  
 FT L -> P (IN REF. 2).  
 FT CONFLICT 295 295  
 FT K -> E (IN REF. 2).  
 FT CONFLICT 309 309  
 FT L -> W (IN REF. 2).  
 FT CONFLICT 352 352  
 FT MISSING (IN REF. 2).  
 FT CONFLICT 365 365  
 FT A -> P (IN REF. 2).  
 FT CONFLICT 612 613  
 FT SS -> EF (IN REF. 5).  
 FT CONFLICT 711 711  
 FT S -> T (IN REF. 2).  
 FT CONFLICT 724 740  
 FT MISSING (IN REF. 2).  
 FT CONFLICT 787 796  
 FT RTRRAPTEP -> PEDIPAPTES (IN REF. 2).  
 FT CONFLICT 804 804  
 FT G -> L (IN REF. 2).  
 FT CONFLICT 814 814  
 FT S -> F (IN REF. 2).  
 FT CONFLICT 817 817  
 FT A -> S (IN REF. 2).  
 FT CONFLICT 889 889  
 FT G -> R (IN REF. 2).  
 FT CONFLICT 1023 1030  
 FT SRSAPPA -> MEAWDHP (IN REF. 3).  
 FT CONFLICT 1034 1034  
 FT A -> AKPVFPFA (IN REF. 2).  
 FT CONFLICT 1894 1894  
 FT K -> T (IN REF. 4).  
 FT CONFLICT 2494 2494  
 FT P -> A (IN REF. 4).  
 SQ SEQUENCE 2517 AA; 274031 MW; F5805C01761258C0 CRC64;

Query Match 32.8%; Score 4147.5; DB 1; Length 2517;  
 Best Local Similarity 40.7%; Pred. No. 1.3e-146;  
 Matches 1099; Conservative 341; Mismatches 785; Indels 475; Gaps 108;

Qy 17 QSRYPHSHVQTFNTRHQEFVADYDYSRHSHLEYSQASQLLQOQQOQLRRPSLLSEFH 76  
 Db 16 EPRYPHSHVQTFNTRHQEFVADYDYSRHSHLEYSQASQLLQOQQOQLRRPSLLSEFH 73  
 Qy 77 PGSDRPQE--RRTSYEPHPGSPVDHDSLESKRPLRQVSDSHFQVSAALPLVHPLP 134  
 Db 74 PGNERSQELHLPESHVLPGLKSEMEFIESKRPLRLELPD-----PLLRSP 122  
 Qy 135 ---EGLRA-SADAKDPAGKHEAPSPISGQCGDDQNASPKLSKEELIQSDRVDR 190  
 Db 123 LLATQPAAGEDLTADRSLTKLE--PVSPGPPHPTDPELELVPLRSLKEELIQSDRVDR 181  
 Qy 191 ETAKVEQILKLLKQOQLEAEAKPEPEKVPSPPPVEQKHSRVQIYYDENRKAEEA 250  
 Db 182 EITMVEQIISLKKQOQLEAEAKPEPEKVPSPPPLESKHSRVQIYYDENRKAEEA 241  
 Qy 251 HKIFEGLPKVELPLYNQPSDTKVYHENIKTNQVMRKLLILFFKRRNHARKQEQKICQR 310  
 Db 242 HRILEGLPQVELPLYNQPSDTROYHENIKINQAMRKKLILYFKRRNHARKQEQKICQR 301  
 Qy 311 YDQLMEAEKKVDRIENNRKAKESKTRYEYEQFPIRQORQOERFQ-RVGORGAGL 369  
 Db 302 YDQLMEAEKKVRIENNRKAKESKTRYEYEQFPIRQORQOERFQ-RVGORGAGL 361  
 Qy 370 SATATARSEHSEIIDLSEQENNEKQRLSVIPPMFMDEORRVKFINNGLMEDPMK 429  
 Db 362 SMSAARSEHSEIIDLSEQENNEKQRLSVIPPMFMDEORRVKFINNGLMEDPMK 421

Qy 430 VYKROPMNVWTDHEKEIFKDKFIOHPKNFGLIASYLKRSKSPDCVLYLYLTCKNENYKA 489  
 Db 422 VYKROQVNMWSEQKETFRKFWQHPKFNGLIASFLKRTVAECVLYLYLTCKNENYKS 481  
 Qy 490 LVRRNYGRRGNQOIARPSQEEKVEEKE--DKAEKTEKKEEKKDEEKDEKDSKEN 547  
 Db 482 LVRRSY--RRRKSQOQQOQQOQQOQQOQPMRPSQEEKDEKEKEAEKEEKEPEVEN 540  
 Qy 548 TKE---KDKIDGTA-EETEEDEQATPRGRKTKANSOGRKRGKITRSMWNEAAAAA 603  
 Db 541 DKEDLLKEKTDGTDGDNDEKAVASKRKTANSOGRKRGKITRSMWNEANSEEAITPQQ 600  
 Qy 604 TEEPPPLPPPEPTEPSTETSWTEEMEVAKGLVEHGRNMAAIAKMWGTSEAOCK 663  
 Db 601 S-----AELASMEINNESSRWTEEMETAKGLLEHGRNWSAIARMVSKTVSOCK 650  
 Qy 664 NFYFNKRRHNDLNLQOQKQKTSRKPREEDVDSQESVASTVSA---QEDDEDIAS-- 717  
 Db 651 NFYFNKRRQNLDELQOQKLMKERNARRKKKAPAAASEEAAFPVVEDEMEASGV 710  
 Qy 718 --NEENPEDSE-----VEAVK----- 732  
 Db 711 SGNEEMVEEAEALHASGNEVPRGECGSPATVNNSSDTPESPHTEAAKDTQNGPKPP 770  
 Qy 733 -----PSEDSPENATSRGNTPEPAVELE---PTTETAPSTSPSLAVSTKPADES 779  
 Db 771 ATLGADGPPGPPPTPPRTSRAPTEPTPASEATGAPTPPAP-PPSP-APPVVPKEEKE 828  
 Qy 780 VETQVNDISAEATQMDVDQOEHSAEESGVCDDPPPATK--ADSDVDVVRVVRPENHASVE 837  
 Db 829 EETAAAPPV-----EBGEQKPPAAEELAVDTGKAEPEVKSECTEAE 871  
 Qy 838 GDNTERKDLRA-----SEKVEPRDEDLVVAQINAOORPEQSDNDSSATCSADE- 887  
 Db 872 EGPAGKDAEAAEAETAGALKAEKKEGSGRATTAKSSGA----PQ-DSDSATCSADEV 926  
 Qy 888 -DVGEPEPQRORFMDSKPSLLNPTGSLV--SSPLKPNPLDLPOLOHRAAVIPPMVSCPT 945  
 Db 927 DEAGGDKNRLSP--RPSLLTPTGDPDRANASQK--PLDLKQLKQRAAIPPI----- 976  
 Qy 946 CNIPIGTPVSGYALYQRIKAMHESALLE-----EQRQRQEQIDLECRSSTSP 993  
 Db 977 -----QVTKVHEPPREDAAPTKAPPAPPQPNLQLOPESDAPQPGSSP 1019  
 Qy 994 CGTSKSP-----NRE-----W-----EVLOPAPH-----QLITNLP 1019  
 Db 1020 RGRSRPAPPADKAEFAAAEAKLPDPPCWTSGLPFPVPPREVIRKASHPADPSAFSTAP 1079  
 Qy 1020 EGVRLP-----TTTRTPPPPLIPSSKTTVAASEK-PSFI--MGSGISQGTPTGTYLTSNQ 1071  
 Db 1080 PGHPLPLGLDHTARPVLPRPTISNPPPLISSAKHPSVLERQIGAISOG-----MSVQLH 1134  
 Qy 1072 ASYQETPKPSVGSISGLPRQOESAKSATLPYIKQEFSPRSQNSQPEGLLV-RAQHEG 1130  
 Db 1135 VPYSEHAKAP-VGPVTMGLPLPMDPKLAPESGVKQEQLSRPGQAGPPESLGVPTAQAS 1193  
 Qy 1131 VVRGTA-GAIOEGSTTRCTPTSKISVESIPSLRGSITGTPALPQTGIPTEALVKGFSR 1189  
 Db 1194 VLRGTALGVPGGSTIGIPSTRVPSDAITRGSIHTGTPA-----DVLTKGTITR 1245  
 Qy 1190 MPIDSSP--EKGREAAKSHVIYEGKSHLSYDNIKNA-----REGTRSPRTAHS 1242  
 Db 1246 IIGEDSPRLDRGREDSLPKGHVIEGKKGHVLSEGGMSVTQCSKEDGRSSGPPHETA 1305  
 Qy 1243 L-KKSYESVEGNIQOGMSMRSPVSAPLEGLICRALP--RGSPHDLKERTVLSGSIHQ 1299  
 Db 1306 APKTYDMGREGVRGATS-----SASIEGLMGAIRPHERSPH-HLKEQHHSITQSG 1358  
 Qy 1300 TPRTTTFSEGL-KYPKQIKRES-----PPIRAFEGAI-----TKGK-YDG-ITFI 1344  
 Db 1359 IPRSYVEAQEDYLREAKLLKREGTPPPPPSRDLTEYKTAQLGPKLKAHAGLVAIV 1418  
 Qy 1345 KEMGRSITHEIPRODILTQESRKTPEVQSTRPIIEGSIQGTPIKFDNN-SQSAIKHNV 1403





Db 1127 ---LGAISQOQMSVOLRVPHS-----EHAKAPMGPIUMLGLPLAVDPKKLGT----- 1169  
Qy 1110 FSPRSQNSQEGLLVRAHQEGVGRGTAGTAQEGSITRTGTSTKISVESIPSLRGSITQGT 1169  
Db 1170 ---ALGATSQSGITKGLPSTRAA---DGPSYRGSITHT 1202  
Qy 1170 PALDQGTIPTEALVKGSIKRMPIEDSP---EKGREEAASGHVYIEGKSGHLSYDNKN 1227  
Db 1203 PA-----DVLKGTISIRVIGSDSPSLDRAREDLPKHGVYIEGKSGHLSYEGGMS 1254  
Qy 1228 A-----REGTRSPRTAHEISL-KESYSEVGENIKQGMSPRESVAPLEGILICALP-RG 1280  
Db 1255 VSQCKEDGRSSGPPHETAPKPTDMMGRVGRITV-----SASIEGLMGRAPTEQH 1308  
Qy 1281 SPHDLKERTVLGSIQMGTPRATTESEFGL-KYPOIKRES-----PPIRAFEGAITK 1334  
Db 1309 SPH--LKEQHIIIRGSIQIGIPRSYVEAQEDYLREKALLREGTPPPPPPRDLTEIY-K 1365  
Qy 1335 GKPYDG-----ITTIKEMGRSIIHEIPQDILTOESRKTTEVVOVOSTRIIEGS 1381  
Db 1366 PRPLDPLGLKLPKTHEGVVATKEAGRSIHEIPREL-----RRTPELPLAPRLKEGS 1420  
Qy 1382 ISQGTPIKEDNNSQSAL-KHNKSLITGSKLSRGMPLEIVPENIKVVERGYEDVKA 1440  
Db 1421 ITQGTPLKIDSGAPSTGKHKHVSIIIGSPRPPPALHLDIMAD-ARALERACYE----- 1475  
Qy 1441 GETVRSRHTSVSSGSPVLRS---TLHEAKQAQLSPGIDYDTSARRTPVSYONTMSRGP 1497  
Db 1476 -ESLKSRSSTSGAGGSITRCAPVVVVELGKPROSLTYEDHGA-----PFTSHLPGSP 1529  
Qy 1498 MNNTSDVTIPP-----NKSTNHERKSTLTPTQRESIPAKSPVPGVDVPSVH--SPFDPHH 1551  
Db 1530 VTTREPTPLQEGSLLSKASQDRKLTSTPRE-----TAKSPHSTVPEHHPIPSYEHLL 1585  
Qy 1552 RGSTAGEVYWSHLPTQLDP-AMPHRALDPAAYALFQRLSPTPPGYPSQYLY----- 1604  
Db 1586 RGVTVGLRGHIFLAFDPTISIPGIPLEAAAAAYILPRHLAPSPTPHLYPPYLIRGYP 1645  
Qy 1605 ---AMENTROTILNDYITSOQMVLN-----RPDVARGLSPREQPLGLPYPA-TRGIID 1654  
Db 1646 DTALEN-RTIINDYITSOQHINAAASMAQADMLRGLSPRESSLALNYAAGPRGIID 1704  
Qy 1655 LTNMPP-TILVPHPGGTSTPMDRITVPGTITFPPRPYNSASMSGPHPLA-AAASA 1712  
Db 1705 LSQVPHLPVLVPPPGTAPATADRILAYLPTAPPFPSSK-HSSSPLSPGPGTHLAKPTATS 1763  
Qy 1713 ERERERERERERETAASSDLYL---RPGSEQ-----PGRPGSHGYVRSPPSVRT 1763  
Db 1764 SSERERERERERKSLTSTTVEHAPITWRPGTEQSSGAGSSRPASHTHQHSPI-SPRT 1822  
Qy 1764 QETMLQORPSVFOQTNGTSTVITPLDPTAQLRIMPLPAGGPPISQGLPASRYNTAADALAA 1823  
Db 1823 QDA-LQORPSVLHNTSMKGVTVSEP-----GTPTV-----LW 1855  
Qy 1824 LVDAASAPQMDVSKTESKESHAARLEENLRSAAYSEQOQLBQKTEV---EKRSVQC 1880  
Db 1856 ARSTSTSPVRPAATFPPATH--CPGLGTLEGVYPTLMEPVLPLPKTSRVARPERARVDA 1913  
Qy 1881 LYTSSAFPSGKQPHSHSVVSEACKDGPPKSYREELTRGTITTAANFIDVITRQ 1940  
Db 1914 GHAFLTTPPGR-EPASSPSKSSPRLAPFSSH-----TAIART 1952  
Qy 1941 IASDKDARERSSQSSDSSLSLSHRYTSPS-----DAIEVIS 1977  
Db 1953 PAKNLAPHASDPDPATPSADLHRETKQKPSIQLELRLSLYHSGAGYSPGCVPEIS 2012  
Qy 1978 PASSPAPPQEKLTQYQPEVVKNAQENDPTQRYEGP-----LHHYRP---QOESPS 2025  
Db 2013 PVSSPSLTHDKGLSKPLEELESKSHLEGELRHKQPGMKLSAEAAHLPHRLPSPQSPSS 2072  
Qy 2026 PQOQLPSSQAEGMGQVPRHRLITLADHICQITQDFARNQVSSQTPQO---PPTSTFQ 2082  
Db 2073 PLLQATPGIRGH-----QRVVTLAQHISEVITQDTRHH-----PQQLSGPLAPLY 2119

Qy 2083 NSPSALYSTPV---RTKTSNRYSPESQAQSVHHQRPGRSVSPENLVKSRGSRFGKSPER 2139  
Db 2120 SFPCA---SCPVLDLRRPPSDLYLPPD-----HCTP-ARGSPH-----SEGKRSPEPSK 2166  
Qy 2140 SHV---SSEPYEPISPQ---VPVYHEKQDSLILLSQSGABEABQORNDARSFGSISYLSF 2194  
Db 2167 TSVLGSSDATEIPSPPEGMTEPGHARSTAYPLL-YRDEQGEPRMGLESFGNTSQPPTF 2225  
Qy 2195 FTKL-ENTSPMVKSKQEIFRKLNSGGGSDMAAOPGTGIFNLPAVTTSGSVSSRGHS 2253  
Db 2226 FSKUTESANSAMVSKKQKQKINLKNHNEPEYINIGPTEIFNPAITGAGLMTCSRQA 2285  
Qy 2254 FADPAS-NLGLIEDIIRKALMGSDFKVEDHGVMSQPMGV---VPGTANTSV-----VT 2303  
Db 2286 VQEHASTNMGLEAIIRKALMGKYQWEE-----PPPLGANAFNPLNASILPAAAMPIT 2339  
Qy 2304 SGETRREGDPSHSGVCKPKLISKNSRKSPIQCGYLGTERPSSSVSVHSEGYH 2363  
Db 2340 TADGRSDHALTSPGGG--KAKVSGRPSSRKAKSPAP--GLASGDRPPSVSVHSEGCN 2395  
Qy 2364 ROTP-GMAWEDRPSSTGCTOPYNPLTMRM---LSSTPPTPIACAPSAYNOAAPHQON 2417  
Db 2396 RRTPLTNVWEDRSSAGSTFPNPLMLQAGVMASPPPPGLAAGSGPL--AGPH--- 2450  
Qy 2418 RIWEREPAPLISAQYETLSDSD 2439  
Db 2451 HAWDEEPKPLCSQYETLSDSE 2472

RESULT 5  
NCRL\_RAT  
ID NCRL\_RAT STANDARD; PRT; 533 AA.  
AC O9WB5: 070463;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR) (FRAGMENT).  
GN NCOR1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=9931771; PubMed=10441327;  
RA Boutell J.M., Thomas P., Neal J.W., Weston V.J., Duce J., Harper P.S., Jones A.L.;  
RT "Aberrant interactions of transcriptional repressor proteins with the  
RL Hum. Mol. Genet. 8:1647-1655(1999).  
RN [2]  
RP SEQUENCE OF 476-528 FROM N.A.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=99421707; PubMed=10491148;  
RA Schuler M.J., Buehler S., Pette D.;  
RT "Effects of contractile activity and hypothyroidism on nuclear hormone  
RL Eur. J. Biochem. 264:982-988(1999).  
CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE  
CC ABSENCE OF LIGAND (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
CC DOMAINS (ID1 AND ID2).



CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND  
CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES  
CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
CC SPECIFICITY.  
CC  
CC -1- SIMILARITY: CONTAINS 2 CORNR BOX.  
CC  
CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
CC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF124821; AAD32566.1; -  
CC EMBL: AF059311; AAC14567.1; -  
CC PROSITE: PS50090; MYB.3; PARTIAL.  
CC Nuclear protein; Transcription regulation; DNA-binding; Repressor.  
KW NON\_TER 1 1  
FT DOMAIN 48 59 POLY-SER.  
FT DOMAIN 153 157 CORNR BOX OF ID1.  
FT DOMAIN 357 361 CORNR BOX OF ID2.  
FT CONFLICT 484 484 R -> W (IN REF. 2).  
FT CONFLICT 497 497 A -> V (IN REF. 2).  
SQ SEQUENCE 533 AA; 57794 MW; 7DF60F8228227EC2 CRC64;

Query Match 19.7%; Score 2493.5; DB 1; Length 533;  
Best Local Similarity 90.08; Pred. No. 4e-86;  
Matches 484; Conservative 21; Mismatches 26; Indels 7; Gaps 3;

Qy 1905 KDKGPKPKRYEELRTRGTTTAAFNIDVITRQIASDKADRGSSQSSSSLSH 1964  
Db 1 KDKGPKPKRYEELRTRGTTTAAFNIDVITRQIASDKADRGSSQSSSSLSH 60  
Qy 1965 RYETPSDAIEVISPASSAPPOEKLQYQPEVVKRANAENDPTQYEGPLHHYRQOESP 2024  
Db 61 RYETPSDAIEVISPASSAPPOEKLQYQPEVVKRANAENDPTQYEGPLHHYRQOESP 120  
Qy 2025 SPQOQ--LPPSSQAGMGQVPRHRLITLADHICQIITQDFARNQVSSQTPQPPSTTFQ 2082  
Db 121 SPQOQPLPPSSQAGMGQVPRHRLITLADHICQIITQDFARNQV--PSQPSTSTFQ 176  
Qy 2083 NSPSALYSTPVRTTSNRYSPESQAQSVHQRPCSRVSPENLVYDKSRGSRPKSPERSHV 2142  
Db 177 TSPSALSTPVRTKPSRYSPESSQSVLHPRGPRVSPENLVYDKSRGSRPKSPERSHI 236  
Qy 2143 SSEPEYTPSPQVPPVHVEKQDLSLLSQRGAEPAEQRNDARSPGISYLPSPFTKLENTS 2202  
Db 237 PSEYEPISPPQGPVAVHEKQDLSLLSQRGMPEAQSDSPGISYLPYFFTKLESTS 296  
Qy 2203 PMVSKKQEIFRKLNSGGGSDMAAQAQPGTEIFNLPAVTTSGVSRGSHSFADPASNLG 2262  
Db 297 PMVSKKQEIFRKLNSGGGSDMAAQAQPGTEIFNLPAVTTSGAVSRSHSFADPASNLG 356  
Qy 2263 LEDIIRKALMGSDKVEDHGVQMSQWGVVPGTANTSVTSGTRREEDGPPSPHSGVC 2322  
Db 357 LEDIIRKALMGSDKVEDHGVQMSQWGVVPGTANTSVTSGTRREEDGPPSPHSGVC 415  
Qy 2323 KPKLKSNSRKSPIPGQYLGTRPSSVSSVHSGDYHROTPGNAWEDRPSSTGSGTQ 2382  
Db 416 KPKLKSNSRKSPIPGQYLGTRPSSVSSVHSGDYHROTPGNAWEDRPSSTGSGTQ 475  
Qy 2383 FPNPLTMRMLSSPTPTPIACAPSAVNAQAAPHQONRIWEREPALLSAQYETLLSDSD 2440  
Db 476 FPNPLTMRMLSSPTPTPIACAPSAITQAAPHQOSRIWEREPALLSAQYETLLSDSD 533

RESULT 6  
MAPB\_HUMAN

ID MAPB\_HUMAN STANDARD; PRT; 2468 AA.  
AC P46821;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE MICROBULE-ASSOCIATED PROTEIN 1B [CONTAINS: MAP1 LIGHT CHAIN LC1].  
GN MAP1B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95104835; PubMed=7806212;  
RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;  
RT Cloning of human microtubule-associated protein 1B and the  
RL Genomics 22:273-280(1994).  
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.  
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS  
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
CC STABILIZING MICROTUBULES.  
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
CC WITH MAP1A AND MAP1B PROTEINS.  
CC -1- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE  
CC KKEE AND KKEI/V, REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER  
CC REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES  
CC BOTH IN VITRO AND IN VIVO.  
CC -1- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED  
CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH  
CC BOTH MAP1A AND MAP1B.  
CC -1- SIMILARITY: TO NEURAXIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: L06237; AAA18904.1; -  
CC MIM: 157129; -  
CC InterPro: IPR000102; -  
CC Pfam: PF00414; MAP1B\_NEURAXIN; 10.  
CC PROSITE: PS00230; MAP1B\_NEURAXIN; 6.  
KW Microtubules; Repeat; Phosphorylation.  
FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.  
FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
FT KKEE AND KKEI/V REPEATS).  
FT DOMAIN 1869 2074 12 x 17 AA TANDEM REPEATS.  
FT REPEAT 1869 1885 1.  
FT REPEAT 1886 1902 2.  
FT REPEAT 1903 1919 3.  
FT REPEAT 1920 1936 4.  
FT REPEAT 1937 1953 5.  
FT REPEAT 1954 1970 6.  
FT REPEAT 1971 1987 7.  
FT REPEAT 1988 2004 8.  
FT REPEAT 2005 2021 9.  
FT REPEAT 2022 2038 10.  
FT REPEAT 2039 2055 11.  
FT REPEAT 2056 2072 12.  
SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;

Query Match 3.8%; Score 480; DB 1; Length 2468;  
Best Local Similarity 18.7%; Pred. No. 8.8e-11;  
Matches 492; Conservative 375; Mismatches 1009; Indels 762; Gaps 113;

Qy 71 LLSEFHGSDRPOQRRTSYEPFHGSPVDHDSLESKRPRLEQVSDSHFORV---SAAVL 127





QY 2046 HRLI-----TLAD-----HICQIITQDFARNVSSOTPOQPPTSTF 2081  
D 1981 RRLDDISNGYDDSDGHTLGDPSYSYETETKITSFPESEGYSETSTKTRTPDSTY 2040  
QY 2082 -ONSPSALVSTPVRTK-----TSNRYSPESQAQSV-----HHORPSGRVSPE 2122  
D 2041 CYTAEKTRTPQASTSYETSDLCYTAEKKSPEARQDVLCLVSSCEYKHPKTELSPS 2100  
QY 2123 NLVDKSRGSRGKSPERSHVSSEPEVPSPPQVWVHEKQDLSLLLSQGAEPAE----- 2177  
D 2101 FI-----NPNLEWFASEEPEESEK-----LTQSGGAPPPPGKQ 2137  
QY 2178 --QRNDARSPGISYLPFFFTKLENTSMWKSKEIFRKLNSGGGSDMAAAQPGTEI 2235  
D 2138 QGRQDETPTFSVS--ESAPQTSDDVPETEECPSTADANIDSESEIPIIDKTVY 2195  
QY 2236 FNL-----PAVTTSGSVSSRGH---SFADPASNLGLIEDIIRKALMGFDDKVEDHGVMSQP 2289  
D 2196 KHMDDPPAPVQDRSPR-HPDVSMDPEA-LAIBQNLGKALKDKLKEKTK----- 2246  
QY 2290 MGVPVGTANTSVVTSGETRREGDPSHSGGVCPRKLLSKNSRKS--SPIQGGYLG 2347  
D 2247 ---KPGTKTS---SSPVKSDGSKPLAASP-KPAGLKSSDKVSRVASPKKXESVEKA 2299  
QY 2348 ERPSVSSVHS-----EGDYHRTQPCWAWEDRPSSTGTSQFFVNPMTMLSLSTPTPI 2401  
D 2300 AKPTTTPVKAARGEKDKETKNANASAKSAXT-ATAGPGTKTKSSAVPGLPV 2356  
RESULT 7  
MAPA\_RAT  
ID MAPA\_RAT STANDARD; PRT; 2774 AA.  
AC P34926;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2].  
GN MAP1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain.  
RX MEDLINE=92355629; PubMed=1379599;  
RA Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;  
RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in  
one messenger RNA.";  
RL J. Biol. Chem. 267:16561-16566(1992).  
CC -!- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN THE FILAMENTOUS  
CC -!- CROSS-BRIDGING BETWEEN MICROTUBULES AND OTHER SKELETAL ELEMENTS.  
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
CC WITH MAP1A AND MAP1B PROTEINS.  
CC -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT  
CC APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE  
CC THEIR MORPHOLOGY.  
CC -!- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE  
CC FOR THE BINDING OF MAP1A TO MICROTUBULES.  
CC -!- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.  
CC -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED  
CC FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH  
CC BOTH MAP1A AND MAP1B.  
CC -!- SIMILARITY: TO MAP1B.  
CC  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
DR EMBL: M83196; AAB48069.1; -  
DR PIR: A43359; A43359.  
KW Microtubules; Repeat; Phosphorylation.  
FT CHAIN 22465 2774 MAP1 LIGHT CHAIN LC2.  
FT DOMAIN 309 496 LYS-RICH (BASIC).  
FT DOMAIN 336 541 11 X 3 AA REPEATS OF K-K-[DE].  
FT REPEAT 336 338 1.  
FT REPEAT 415 417 2.  
FT REPEAT 420 422 3.  
FT REPEAT 424 426 4.  
FT REPEAT 427 429 5.  
FT REPEAT 431 433 6.  
FT REPEAT 436 438 7.  
FT REPEAT 440 442 8.  
FT REPEAT 444 446 9.  
FT REPEAT 449 451 10.  
FT REPEAT 539 541 11.  
SQ SEQUENCE 2774 AA; 299536 MW; 3DEF74427BA9D7D7 CRC64;

Query Match 3.7%; Score 465.5; DB 1; Length 2774;  
Best Local Similarity 19.1%; Pred. No. 3.5e-10;  
Matches 570; Conservative 350; Mismatches 1012; Indels 1053; Gaps 140;

QY 96 PSPVDHDSLEKRPRLQVSDSHFQRYSAAVLPLVHPLPEGLRASADAKDPAGGKHEA 155  
D 18 PSPFD-----LLEPPTSGGFLKLSK---PCYIFP-----GGRGDS 50  
QY 156 PSSPISGQCGDQNASPSKLSKEELQSDMRVDRE-IAKVEQQILK-----LKKKQOOL 209  
D 51 ALFAVNGENILVD--GGDRKSCFWKLVHRLDRIDSVLLTHIGADNLPGINGLLQKVAEL 109  
QY 210 EEEAAK-----PPE-----PEKVPSPPPVQKHRSIVQIYDENRKAEE 248  
D 110 EEEQSSSSYSDWVKNLISPELVGVFVFNVDKRLPDAARAKRSI-----E 157  
QY 249 EA-----HKIPEGIGPKVELPLYNQPSDT-----KVYH-----ENIKTNVMRKKLILF 292  
D 158 EACITLQH--LNLRLGIAOE-PLYRVVSNITIEPLTLFHKMGVGRDLMYVLPNPKDSKEMQF 214  
QY 293 FKRR--NHARKQ-----REQICQRYDQLEMA--W-----EKKVDRI---ENNPRR 331  
D 215 LMQKWAGNSKAKTGIVLANGKEAELISVPLYTSITLVVWLPANPTEKIVRVLPFGNAPQN 274  
QY 332 KAKE--SKTRYEYKQPP-----EIRKQEQOERFORVGOAGAGLS 370  
D 275 KILEGLEKLRHLDLRYVPVATQKDLAAGAVPANLKPSPKIKHRADSKESLKAAPTAV--- 331  
QY 371 ATARSHEISEIIDGLSEQEN-----NEQMQRQLSVIPPMFMFAEQRRVFINMNG 422  
D 332 SKLAKREEVLEE---GAKEARSELAKELAKTEKRAKESEKPP-----EKPSRSVRG 382  
QY 423 LMEDPMKYKDRQFMNWTDEHEIKFDKFIQHPKFNGLIASYLERKSVPCVLYYYLTK 482  
D 383 ESSALAKERR-----LIKDK----- 399  
QY 483 KNEWKALVRNRYKRRGRNQOIARPSQEEKVEEEDAEKTEKKEEKEDEKDEKE 542  
D 400 -----AGKH-LKEKISK-LEEKDKKEKEIKERKELKKEGEEKKDAKK 445  
QY 543 DSK-ENTKEKIDGTAETEERQATPRGKT---ANSOGR---RKGRITR---SMTNE 592  
D 446 DEKRDKTKPEVK-----KLSKPDLPKPTFEVKTLYKAKAPGRVVDKRAARGEKELSS 501  
QY 593 AAAASAAAAATEPPPPPPPISTPEVETSRWTEEMEVAKKGLV----- 641  
D 502 PRTPPAQKGA--PPAAVSGHRELALSSPEDLTQ-DFEELKREERGLLAQRDTGLGEK 557  
QY 642 -----EHRGNWAAI--AKMVGTKSEAQCKNFYFNKRRHNLNLLQOQKQKTSRPR- 691  
D 558 PLPADATEQGHPSAAIQVOTPSGPLEGE-----HVEREKEVVPDPSGDKGSTNRGDS 611

Qy 692 -----EEDVSOCEVASTVAQEDIEDIASNEE--ENPDSRVEAVKPSDEDSPE 739  
Dy 612 GAVEKEKETWEERKORAEALGPENTAREESEAUVKEDVIEKAELEMEETHESDEGE 671  
Qy 740 NATSRG-----NTEPAVELEPTT-----ETAPSTSPSLAVPS----- 771  
Dy 672 ETKAESFYQHTQKALKASPRKALGRDLGFGOKAPEKETATFSL-SLATPAGATEHV 730  
Qy 772 -----TKPAEDESVEQVNDISIAETA-----QMDVDOEHSABEGSVCDPPP---AT 817  
Dy 731 SYIOETIPGSETQETISDEIHEDEDERPAPRFTPTVDLSGPEG-----PGPFRAQ 786  
Qy 818 KADSV-----DVEVRVP-----ENHAKVEGDNTKERDLDRASEKVEP 855  
Dy 787 AADSAPASSSKTYGAPETELTYPPNMVAAPLAEHVSNA-----TSITECDKLSSRATS 842  
Qy 856 RDELVVAQINARPE-----PQSDNDS-----SATCSADED 888  
Dy 843 VAEDQSASLTAPQTEETGKSSLLDVTISIPSSRTEATQGLDYPVSAAGTISPTSSLEED 902  
Qy 889 -----VDEPE-----RORMRPMDSKPSLLNPTGSLVSSP 919  
Dy 903 KGFKSPCEDFSVTGESEKGETVGRGLSGEKA VGEKYEKVVVTSEK--LSQYAAVFCAP 960  
Qy 920 ---LKNPPLDLPOLQHRA-----AVIPPMVSCPTPCNIPITGTPVSGYALYQRIKAM 967  
Dy 961 GHTLPPGEPALGEVEERCLSPDDSTVKMASPP-----PSGPPSAAHTPF----- 1004  
Qy 968 HESALLEQRQRQIDLECHRSSTPCGTSKSPNREVEVLQAPHQILNLPPEGVRPPT 1027  
Dy 1005 HQSPVEDKSEPRDFQEDSWGTEKNSP-GVSKEDSEE-QTVKPGPEEGTSE--EGKGPPTR 1060  
Qy 1028 RPTRPPTP-----LIPSSKTTVASEKPSFIMGSIQGT-PGYLTSNHOAS 1073  
Dy 1061 SPQADMPVSTAGQGTCTIQLPEQKAI VFTGE--AGSNLGAGTLPGEVRYSTEEAT 1118  
Qy 1074 YTOETPKPSVGSISLGLPROQESAK--SATLPYIKQEEFSPRS-----QNSOPEGIL----- 1123  
Dy 1119 EPQKDEVLRFDTQSL-S-PEDAESLSVLVSPDPTKQETPRSPCSLKEQKQPHKDLMPW 1177  
Qy 1124 -----VRAQHEGVVGT-----CAIQEGSITRGT-----PYSK----- 1152  
Dy 1178 SPEDQSLSFSESPKSTLSDISSKQSLSPESLGTQFGLNLGKEERGYPYMKAEEDSCH 1237  
Qy 1153 ISVESIPS-LRGSTQGTTPALPQGTIPEALVKGIS-----PAGTLPFGGSFHSALSVDKXHSGETGPGGHFWTSDSS 1297  
Dy 1238 LAPVSIPEPHRATVSPSTDETPAGTLPFGGSFHSALSVDKXHSGETGPGGHFWTSDSS 1297  
Qy 1189 -----RMPIEDSPPEKREEAASKGHIYIEGKSGHILSYDNK 1226  
Dy 1298 LTKSPESILSPAMEDLAVEWEGKAPGKEKEPELKSETRQKQKQILPEKVA--VVEQDLII 1355  
Qy 1227 NAREGT-----RSPRTAHEISLKSRYESVGNKQKMSMRSPSVAPLEGICRAL 1277  
Dy 1356 HQKDGDALDEENKPGQODKTPQKGRDLDEKDTAAELDKGPEKPKDLREDQO----- 1409  
Qy 1278 PRGSPHSDLKERTVLGSLMOGT-----PRATSEFEGDKYPKOIK-RESPIRAFEGAI 1332  
Dy 1410 QRAGPPAEKKASQORDDLOQTQATEPRDRAQRRORESEKDKSLEULDRTP----- 1461  
Qy 1333 TKGRPYDGITTIKEMGRSIIHPIRQDILTOESR-----KTPBVQOSTRPIIEGSISSQGTPI 1388  
Dy 1462 -----EEKDR-----ILVQEDRAPEHSIPEPTQTD-----APDKGTDD 1496  
Qy 1389 KFDNNSGOSAIKHN-----KSLITGPKSLSRGMPLEIVPENIKYVVERKGYEDVAGETVR 1445  
Dy 1497 K-EQKEASEEKEQVLEQKQWALKEG-----ETLDQEARTEAO-KDETLEDKTK-Q 1545  
Qy 1446 SRHTSVVSSGSPVLSRLTHEAPKAQLSPGIYDDTSA-----RRTPVSYQNTWNRSGSPMNR 1502  
Dy 1546 GQKSSFVEDKTTTSKETVLQOKSAEKADSVEQQDGALEKTRALGLAESPAEGS----- 1599

Qy 1503 SDVTIPPNKSNHERK-----STLTPTQRESIPAK-SPVPG-VDPVWVSHSPDPHHRG--- 1553  
Dy 1600 -----KAREQEKYKWEQODVQGWRETSPTRGEPVGGQKEPVPANEKGSPEQEVRYW 1651  
Qy 1554 -----STAGEVYWSHLPTQLDPAMPFHRALDPAAA--AYLFQRLSPPTGPGVPSQQLYAM 1606  
Dy 1652 RDROITLQDDAYWREL--SCDRKVMFPHELDQCARPRYCEBERESTFDEGDEQEIPL 1709  
Qy 1607 ENT-RQTLINDYITSOQMVNLRPDVARGLSRPOPLGLP---YPATRGIIIDTINMP-- 1659  
Dy 1710 QHTPRSPWTSDFKDFQEPPLPOKGLVEVERKLA--ESPVLPEEEDKLTFRSEFELISPPAS 1767  
Qy 1660 -----PTILVPHPGGTSTPPMDRITYIPGTOITPPP-----RPNYSASMSPG- 1701  
Dy 1768 PPENTGORVPSAPGOESPVPDTESTAPMRNEPTTPSWLAEIPPMVKDRPLPAPLSPAP 1827  
Qy 1702 -----HPTHLAAAASAERERE-----REKEKERERERI 1729  
Dy 1828 APPTPAPEPHTPVFPFMSGLAEYDSVAAVQEGAAELEGGPSPLGKDYRKABGEREGEG 1887  
Qy 1730 AAASDLYLRPGSEQGRPGSGHYVRSPPSVRTQETMLQORPSVFOGTNGTSTVITPLDP 1789  
Dy 1888 ACAPDSSSFSPKVPPEAGESLATRTEQTEPEQREPTPYPDERSFOYADIYEQMLTLGLP 1947  
Qy 1790 TQALRIMPLPAGG--PSISQGLPASRYNTAADALAAALVDAASAQMDVSKTESKHEA 1846  
Dy 1948 ACPTREPPLGASGDWPHLSTKEEAAGCNTSAE-----KETSSPA 1987  
Qy 1847 ARLENLRSRAANVSEOOOLEQKLEVEKRSVQCLYTSSAFPS-----CKPQPHSSVWTS- 1901  
Dy 1988 S--PQNLSQDTPAFS-----YASLAGPAVPPROEPDPPGPNVEPSI 2025  
Qy 1902 -----EAKDKGPP-----PKSRVEEELRTRGKTTITTAANFIDVITIQI 1941  
Dy 2026 TPVAVPPRAPISLSKDLSPPLNGSTVSCSPDRRPPSPKETGRGHWDGNDSDL----- 2079  
Qy 1942 ASDKDARERGSQ-----SDSSSSLSLSHRY-----ETPSDAI 1973  
Dy 2080 --EKGAEOPEKETRSPRPHMPMGHSSLWPEATEAYSSLSSDSHLGSVRPSLDPPASAF 2137  
Qy 1974 EVISPASSPAPPOEKLOTYQPEVVKANQA-----ENDPTROYEGPLHHY 2017  
Dy 2138 GFSS--LOPAPPO-----LPSPAEPRSPAPCGSLAFSGDRALALVPGTPTTRH---DEY 2186  
Qy 2018 RPOESPSPOOOLP--PSSQAEG--MGQVPRTHRLITLADHICQIITQDFARNQVSSOT 2072  
Dy 2187 LEVTKAPSLDSSLQPLSPSPSGGGLLSNLPR----- 2218  
Qy 2073 PQOPPTSTFQNSPALSALVSTPVRTKTSNRYSP-----ESQAQSVHHQRPG 2116  
Dy 2219 --PASPALSEGSSSEATTPISSVAERFPGLAEAAEQSAEGLSGSGKESAHSWLDLTP 2275  
Qy 2117 SRV-----SPENLVDKSRGSRPGKSPERSHVSS 2144  
Dy 2276 SPAPASLIDLAPAPAPAPAPAGPLGDIGDGTLPFCRPECTGELTKKPSFPLSPSGDHEAN 2335  
Qy 2145 EPEY-PTSPPO-VPVWHEKQDLSLLLS--ORGAEPAEQRNDARSPCGSIYLSFFFTKLENT 2201  
Dy 2336 PGGETSLNPPGFTVATAEKEEAPAWERSWPBGAEARSRPD-----TLLSE 2385  
Qy 2202 SPVMSKKQEIFRKLNSGGGSDMAAAQPGTEIFNLPAVTTSVGSVSRGHSFA--DPASN 2260  
Dy 2386 QPLRPGK-----SSGG-----PPCSILSSEVEAGPOGCATDPRPH 2419  
Qy 2261 LGLEDITRKALMGSDDKVEDHGVMSQPMGVPGTANTSVVTSGET-----RRREGD 2313  
Dy 2420 CG-----ELSPSF-----LNPLP--PPSTDSDLSSTEARELAGKGRRRVR 2459  
Qy 2314 PSHSGGVC-----KPKLISKNSRKS KSPIGO-----GYLGE 2348  
Dy 2460 PGA-TGGCPMADETPPTTSASDSGSSSDVPPPETECPSTAEALDSDEGDFLVD 2518  
Qy 2349 RPSVSSVHSEGDYHRQTPGWAWEDRPSSTGTSQTPPNPLTMRMLSSPTPIAC----- 2403

```
Db 2519 KAGVSGTHHPRGHPPTPLDPPRPS-----APSAVNOAAPHQONRI-WEREPAP 2426
Qy 2404 -----PPRDPVCMADPE 2558
Db 2559 GLSSGSRVERLREKGRGRRAPGRAPKASPARRLDIRKRSPTP 2603
Qy 2404 -----PPRDPVCMADPE 2558

RESULT 8
MAPB_MOUSE
ID MAPB_MOUSE STANDARD; PRT; 2464 AA.
AC P14873;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MICROTUBULE-ASSOCIATED PROTEIN 1B (MAP1.2) (MAP1(X)) [CONTAINS: MAP1
DE LIGHT CHAIN LC1].
GN MAP1B OR MTAP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
RX MEDLINE=90094539; PubMed=2480963;
RA Noble M., Lewis S.A., Cowan N.J.;
RT "the microtubule binding domain of microtubule-associated protein
RT MAP1B contains a repeated sequence motif unrelated to that of MAP2
RT and tau.";
RL J. Cell Biol. 109:3367-3376(1989).
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -1- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE
CC KEEE AND KKEI/V. REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER
CC REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES
CC BOTH IN VITRO AND IN VIVO.
CC -1- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC OF MAP1B.
CC -1- SIMILARITY: TO NEURAXIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X51396; CAA35761.1;
CC PIR: S07549; QRMSP1.
CC MGD: MGI:97179; Mtap5.
CC InterPro: IPR000102;
CC Pfam: PF00414; MAP1B_neuraxin; 7.
CC PROSITE: PS00230; MAP1B_NEURAXIN; 7.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.
FT DOMAIN 589 787
FT
FT DOMAIN 1865 2068
FT REPEAT 1865 1881
FT REPEAT 1882 1898
FT REPEAT 1899 1915
FT REPEAT 1916 1932
FT REPEAT 1933 1949
```

```
FT REPEAT 1950 1966
FT REPEAT 1967 1983
FT REPEAT 1984 2000
FT REPEAT 2001 2017
FT REPEAT 2018 2034
FT REPEAT 2035 2051
FT REPEAT 2052 2068
SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDBDA87 CRC64;

Query Match 3.5%; Score 447; DB 1; Length 2464;
Best Local Similarity 18.6%; Pred. No. 1.5e-09;
Matches 490; Conservative 375; Mismatches 957; Indels 812; Gaps 116;

Qy 71 LLSFPHGSDRRPQRRRTSYEPFHGCPSPVDHDSLESRRPRLEQVSDSHFQRV---SAAVL 127
Db 174 LLSTTHPAN-----KASLTLCFCP-----EEDGWNNSNLDNRHLQDFINIKLSASIL 220
Qy 128 PLVHPLPEGLRASADAKDPAGFGKHEAPSSPIG-----QPC-----GDD----- 168
Db 221 PEMEGLSEFTEYLSSEVFPs---PFDILEPTSGGFLKLSKPCYIFPGRGDSALFAP 277
Qy 169 -----QNASPSKLSKEELISQMDRVD-----REIAKVEQQ--- 198
Db 278 NGFNMLINGGSRKSCFWKLIRHLDNRVDSILLTHIGDNLPGINSMLQRLAELEERSQ 337
Qy 199 -----TLKLLKKQO-----OLEEAAKPEP 219
Db 338 GSTSNSDMKNLISPDGLGVFLNVPENLKDPENIKMKRSTEEACFTLQYLNLKLSMKPEP 397
Qy 220 -----EKPV-----SPPVBOKHRSIVQIYYDENRKKAEAAHKIFE 255
Db 398 LFRSVGNFIEPVILFQKMGVGLKEMVNLNPKVSKKEMQYFMQOWTGTNKDAE---LIL 453
Qy 256 GLGPKVELPL--YNQPSDTKYVHENIKTNQVMR-----KKLILFFKR 295
Db 454 PNGQVEDIPISYLTSSVSLIWHHPANPAEKIIRVLPFGNSTQYNIILEGLEKLKLDLFLQ 513
Qy 296 RNHARKQ-----REQICQRYDOLMEAWKKVDRIENPRRKAKEST-REYEEK 344
Db 514 PLATQKDLTGVPFPVPPVQVKKLQKQAD-----SRESLKRPATPVASKSVKRESKE 563
Qy 345 QFPPIRQQRQEQRFQVQRGAGLSATIRSEHEISEIIDGLSEQNNENKQRLSVIP 404
Db 564 ETPVTK-TSQVEKTPKVESKEKVL---VKKDQPKVTKESKPSVTEKESVSKESQs----- 614
Qy 405 PMFDAAQRRVKFNMGLMEDPMKVYKDRQFMVWTDHEKEIFDKPIQHPKPNGLIAS 464
Db 615 PVKAEVAEKQA-----TESKPKVTDKVKVKE-----IKT 644
Qy 465 YLERKSVDPDCVLYYYLTKNENYKALVRRNYGKRRGRNQIARPSQEEKVEEKEDKAEK 524
Db 645 KLEEK-----KEKPKKEVKK-----EDKPTLKKDEKPKR 675
Qy 525 TERKEEEKKDEEKDEKSKENTKDKIDGTAETEEREQATPRGKRTANSQGRKGR 584
Db 676 EEVKKEIKKEIKKEERKELKKEVKETPLDKAKKEVKKKEKKEKKEKPKKEIKKISK 735
Qy 585 ITRSWTNEAAAAAASAAAAATEPPPLPPPEPISSTPETSRTWTE-EMEVAKKGLVGH 643
Db 736 DIKSTPQSDTKKPSAL-----KPKVAKKEESTKKEPLAAGKLKDKGKVIKK----- 785
Qy 644 GRNWAALAKWVGTKSEAOCKNFYFNKRRRNLNLLQOHQKQKTSKPREERDVQSCEVA 703
Db 786 GKTEAAATAGTAATAA-----VVAAGTAASGPVKELEAER 824
Qy 704 STVSAQDEDEIASNEENEPDSEV-EAVKPSDESPENATSRGNTPE-----AVELEPTE 758
Db 825 SLNSS--PEDLTDFEELKAEIDVAKDIKPOLIEDEEKLKTKETQCEAVVIQKETEVS 882
Qy 759 TAPSTSPSLAVPSTK-----PADESVETQVNDSI-----SAET----- 792
Db 883 KGSAESPDGITTTEGEGCEQTPPEELPEPVKQGVDDIDKEFDEGAFGESSETGTYEEK 942
```





Qy 1161 \*--LRGSIQ-----GTPALPQTGIPTEALVKGSISRMPIEDSSPE 1198  
|||  
Db 1302 GKYLPGAITSDEHILTXDSSFSSKSPESLXGPAKDIAIKWEDKVPGLKORTSEQKEPE 1361  
Qy 1199 KGREAAKSHV-----IYEKSGHILSYD---NIKAREGTRSPRTAHEISLRKSYESV 1250  
Db 1362 PKDEYLOQKDKTLEHKEVVEPKDTAIYQKDEALHVKN--EAVKQXKALE-QKGRDLRQX 1418  
Qy 1251 EGNIKQGMRESVPASPLEGILICALPRGSPHSDLKERTVLGSGIMOGTTPRATESPED 1310  
Db 1419 DTALEQDKALE-PKDRDLEK-DKALEQKDKIPXKDKAL-----EOKDTALEQDK 1468  
Qy 1311 GLKYPKQIKRESPIRAPEGAITRGK-PYDGITTIKEMGSI-HEIPRODILITQESR--- 1365  
Db 1469 KALEPKDKDLQK-DKLEK-RVLE---QKEKIPKEDKALDQKRVSEHKAPEDTVAEMKDRDLE 1524  
Qy 1366 ---KYPE---VQSTRPIEGISQGTPIK-FDNNSGQSAIKHNKVSILITGPKLSRQMP 1418  
Db 1525 QTKAPEQKHAQEQKQKVSSEKDKQALEKXWALQKDEALEXNQALEENHOTQEQESL 1584  
Qy 1419 PLEIVPENIKVVERGKYEDVKAGETVRSRHTSVVSSGSPVLSRSTLHEAPKAQLSPGIYDD 1478  
Db 1585 VQDKTRPKMLEKSPKVRAME-----EKLEALLEKTKALGLEESLQVE 1630  
Qy 1479 TSARITPVSYQNTMRSGSPMMNRSDVTIPPNKSTNHERKSTLTPTQRESIPAKSPVPGV 1538  
Db 1631 GRAREQEKYV---RGQDVVQEQE-TSPTRREPAGQKE-LAPAWEDTSPQD----- 1679  
Qy 1539 DPVVSHPDPIHNG-----STAGEVYVSHLPTQDPAWPFHRLDPAALYFQRLSPT 1594  
Db 1680 -----NRYWGRDVSLEQDITYWELSCERKVPWPHLEDVQGARPHYTEERSTFL 1730  
Qy 1595 PGYQSOYLYAMEN--TRQITLNDYITSOQMVNLRLDPVARGLSPREOPLGLP---YPA 1648  
Db 1731 DEGPDEQEVPLRHRHATRSWASDFKDFQESSPOKGLVEVERWLA--ESPVGLPPEEEDKL 1788  
Qy 1649 TRG---IIDLNMPTTI-----LVPHPGGTSTPP---MDRITYIPG 1683  
Db 1789 TRSPFEIISPPASPMVMGQVSPAGQESPIPDPKLMPHMKNEPTTISWLADIPPPWPK 1848  
Qy 1684 TQITFPFRPYNSASMSGHPH-----HIAA-----AASERERERE----- 1719  
Db 1849 DR-PLPPAPLSPA---PGPPTPAPESHPTAPFWSGTPEYDSVVAQVGAEELEGPPYSP 1904  
Qy 1720 -----REKERERERTAAASDLRLPGSEQGRPGSHGV--RSPSPSVRTQETMLQORP 1772  
Db 1905 LGKDYKVKYGEREEGREAEDPKSHSKVPEARKSHATTEPQTEPQREPTPYPDERS 1964  
Qy 1773 SVFQGTNGTSVITPLDPTAQLRIMPLPAGG---PSISQGLPASRYNTAAD----- 1819  
Db 1965 FQYADIYEQMMLTGLGPACPTREPPLGAAGDWPCLSTKEAAGRNTSAEKELSSPISPK 2024  
Qy 1820 -----ALAAVLDAAA-----SAPOMQVSKYKE-----SKHEARLEENLR-- 1855  
Db 2025 SLQSDTPTFSYAALAGTVPVPRPEPGSMESPLTPVAPVPRAPILSKGSPPLNGILSC 2084  
Qy 1856 ---RRAVSE-----QQOLEKTLVEKRSVQCLYTSSAFPSGKP----- 1892  
Db 2085 SPDRSPSPKESGRSHWDDTSDSELEKGAQEAKEAQSPPPHPIPMGSPPTLWPTEA 2144  
Qy 1893 -----PHSSVYVSEA-GKDK----- 1907  
Db 2145 HVSPLXSHLXARPSLDFPASAFGFSLEXAPQXPSPAPRSPACGSLAFSGDALAL 2204  
Qy 1908 --GPPPKRYEEELTRGKTTITANFTDVIITQIASDKDARGGSSSSSSLSHR 1965  
Db 2205 APGPPTTRYDEYLE-----VTRKPSLDSLLPOLPSPSPGXPLLSN-- 2246  
Qy 1966 YETPSDALEVISPASSAPPOEKL---QTYQPEVVKANO--AENDPTROYEGPLHHYRP 2019  
Db 2247 --LPRPASPLSEGSSEATPTVSSVAERFSPLEAAEQSGELDPGNE---FAAHLXW 2301

Qy 2020 QOESPSPQOOLPPSSQAEGMGQVPRTHRLITLADHI--COITODFARNQVSS-QTPQOP 2076  
|||  
Db 2302 DLTPLSP--APPASLDLALAPAPSLPG--DMGDGILPCHLFCSEAAATEKPSFPQVPSD 2356  
Qy 2077 -----PTSTFQNSPALYSTVPRKTSNRY--PESQAQSVHHQRGSRVS 2120  
Db 2357 CAANGPTETSPNP-----XPAPAKAENEAAAXPAWEGAMPEGAERS---SRPDTXLS 2408  
Qy 2121 PENLVKDSRGRSPKSPERSHVSSEPY-----EP-----ISPQVPPVHKKQ--- 2162  
Db 2409 PEQPCVPCGGSGPPSSASPEVEAGPQCXYTEPRHGRGELSPFLNPLPPSIDDRDLST 2468  
Qy 2163 DLSLLSQRGAEPQARNDAKSPGISYLPSTFFTKLENTSPMVKSKKOEIFRKLNSGGG 2222  
Db 2469 EEVRLVGRGG-----RRRVGPGGTG-XPXPVT--DETPTPTAS-----DSGSS 2509  
Qy 2223 DSDMAAAGPCTEINFNLPAVTTSGSVSSRGHSFADPASNLGLEDIIRKALMGSFDDKVEDH 2282  
Db 2510 QSD-SDVPPETE--ECPSITAEALDSDGDFLVPDXG-----GVSGTHHPRPGHD 2559  
Qy 2283 GVMWSOPMGVPGTANTSVVTTGETRREEDGDPSPHSGVC--KPKLSKNSRKSPP 2340  
Db 2560 PPPLPQP-----DXRPSPPRPDVCMDPGLSESGR----- 2591  
Qy 2341 GQYLGTERPSVSVHSEGDYHROTQGMWEDRPSSTGTFQYFNPLTMRMLSTPTPT 2400  
Db 2592 -----XERLRKKEV--QGRVGRAPG---KDKPVS-----PXRLKLR--GKRSPTP 2632  
Qy 2401 -----TACAP-----SAVNOAAHQ 2416  
Db 2633 GKGSXDRVSRXPXRSRXTSQVTPAE 2659  
RESULT 10  
APC\_MOUSE  
ID APC\_MOUSE STANDARD; PRT; 2845 AA.  
AC Q61315; Q62044;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ADENOMATOUS POLIPOSIS COLI PROTEIN (APC PROTEIN) (MAPC).  
GN APC.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RC SEQUENCE FROM N.A. (ISOFORM 1 AND 2), AND VARIANTS.  
RP STRAIN=C57BL/6J, AND CAST/EI; TISSUE=Brain;  
RX MEDLINE=92263101; PubMed=1350108;  
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,  
RA Luongo C., Gould K.A., Dove W.F.;  
RT "Multiple intestinal neoplasia caused by a mutation in the murine  
RL homolog of the APC gene.";  
RL Science 256:668-670(1992).  
RN [2]  
RP ERRATUM.  
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,  
RA Luongo C., Gould K.A., Dove W.F.;  
RL Science 256:1114-1114(1992).  
RN [3]  
RC SEQUENCE OF 1-45 FROM N.A.  
RP STRAIN=BALB/C; TISSUE=Liver;  
RA Dicker F., Lambertz S., Reitmaier A., Ballhausen W.G.;  
RT "The murine APC gene: alternative splicing of 5' untranslated  
RL region segments.";  
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP ALTERNATIVE SPLICING.  
RX MEDLINE=94061824; PubMed=9242607;  
RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;  
RT "APC gene messenger RNA: novel isoforms that lack exon 7.";  
RL Cancer Res. 53:5589-5591(1993).

-1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE ALLOWS THE DOWNREGULATION OF CYTOPLASMIC BETA-CATENIN (BY SIMILARITY).

-1- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATENINS (BY SIMILARITY).

-1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE PRODUCED BY ALTERNATIVE SPLICING.

-1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART, LUNG, BRAIN, STOMACH, INTESTINE, TESTIS AND OVARY.

-1- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).

-1- SIMILARITY: CONTAINS 7 ARM REPEATS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M88127; AAB59632.1; -;  
EMBL: U02937; AAA03443.1; -;  
HSP: Q02248; 2BCT.  
MGD: MGI:88039; Apc.  
InterPro: IPR000225; -;  
Pfam: PF00514; Armadillo\_seg; 4.  
PROSITE: PS50176; ARM\_REPEAT; 1.  
Anti-oncogene; Phosphorylation; Alternative splicing; Repeat;  
Coiled coil.  
DOMAIN 1 61 COILED COIL (POTENTIAL).  
DOMAIN 125 245 COILED COIL (POTENTIAL).  
DOMAIN 1 728 LEU-RICH.  
REPEAT 451 493 ARM 1.  
REPEAT 503 545 ARM 2.  
REPEAT 546 589 ARM 3.  
REPEAT 637 681 ARM 4.  
REPEAT 682 723 ARM 5.  
REPEAT 724 765 ARM 6.  
DOMAIN 739 2834 SER-RICH.  
DOMAIN 1130 1156 ASP/GLD-RICH (ACIDIC).  
DOMAIN 1556 1575 ASP/GLD-RICH (ACIDIC).  
DOMAIN 1864 1891 HIGHLY CHARGED.  
VARSPPLIC 243 276 MISSING (IN ISOFORM 2 AND ISOFORM 4).  
VARSPPLIC 310 410 MISSING (IN ISOFORM 3 AND ISOFORM 4).  
VARIANT 120 120 T -> A (IN STRAIN CAST/EI).  
VARIANT 493 493 V -> I (IN STRAIN CAST/EI).  
VARIANT 797 797 Y -> F (IN STRAIN CAST/EI).  
VARIANT 1330 1330 A -> T (IN STRAIN CAST/EI).  
VARIANT 1618 1618 A -> S (IN STRAIN CAST/EI).  
VARIANT 2294 2294 G -> A (IN STRAIN CAST/EI).  
VARIANT 2496 2496 H -> Q (IN STRAIN CAST/EI).  
VARIANT 2523 2523 T -> A (IN STRAIN CAST/EI).  
VARIANT 2813 2813 T -> S (IN STRAIN CAST/EI).  
SEQUENCE 2845 AA; 311086 MW; 145CA73CF570A499 CRC64;

Query Match 3.58; Score 437; DB 1; Length 2845;  
Best Local Similarity 19.4%; Pred. No. 4.le-09;  
Matches 423; Conservative 271; Mismatches 831; Indels 658; Gaps 91;

QY 347 PEIRKQROQR-RFQVQRCAGLSATIRSEH-----EISEIDGL 387  
DB 979 PSVESYEDDESKFCSGYQYADLAHKIHSANHDDNDGELDTPTINSLKYSDEQLNSGR 1038  
QY 388 SEQENNEKQRLSYIPPMFDEAQRVKYFINMGLMDPMKVYKDRQFMNVWTDHEKEI 447  
DB 1039 QSPSONERWARPKHVEIDEIKQNEQARSONTS-----YPVYSEN-----ID----- 1081  
QY 448 FKDKFIQPKNFGLIASYLERKSPVDCVLY-----YLTKKENYKALV 491  
DB 1082 --DKHLKQPHFG-----QOECVSPYRSRGTSGSETNRMGSSHAINQNVNQLCQ 1129

QY 492 RRNYGKRRGRNQIARISQREKVEEKEDKAEKTEKEERKKDEEKKDE----- 540  
DB 1130 EDDYEDDKPTNYS-ERYSEEEQHEEERPTNYSIKYNEEHVDDPIDYSLKATYATDISS 1188  
QY 541 -----KDSKENTKEKDKIDTAETETEEBQATPRGRKTANSQGRKKRGRITR----- 587  
DB 1189 SQKPSFSFKNSAQSITK-PEHLSPSSSENTAVPPSNAKRONQLRPSSAQRNQGTGKTC 1247  
QY 588 -----SMTNEAAAAA----- 602  
DB 1248 KVPISQNETIOTYCVEDTPICFSRCSSLSLSSADDEIGCQDTTQEDASANTLQTAEYKE 1307  
QY 603 -----ATEEPPPLPPPEPISTETPETSRTTEEMEVAKKGL-VEHGRNWAATAKMWGT 656  
DB 1308 NDVTRSEDPAETEPVAVSONARAP-----SRLOASGLSESTRHNKAVEFFSSGA 1357  
QY 657 KSEAQCKNFYENYRRHNLNLLAQHKQKTSRKPREE-----RDVQCESVASTVSQAE 710  
DB 1358 KSPS-----KSGAQTPKSPPEHYVQETPLVFSRSTSV-SSLDSFE 1396  
QY 711 DEDIEASNEENPEDSEVEAVKPS--DSPENA--TSRGNTPEPAVELEPTTETAPSTSPS 766  
DB 1397 SRSTASSVQSEPCSGMVSGIISPSDLPDPGQTMPPPSRSKTPPP---PPQTVQAKREVPK 1453  
QY 767 LAVPSTPAEDESVEVTOVNDISIAETAQMDVDQOEHSAEE-----GSVC-----DP 813  
DB 1454 SKVPAAEKRESGPKQTAVNAAVQ-RVQVLPDVLILLHFAETESTPDGFCSSLSALSILDE 1512  
QY 814 PPATKADSDVDYEVRY--PENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQOINARQP 871  
DB 1513 PFIQK---DVELRIMPVQ-----ENDNGNETESEQEESNEQDKV-----EKP 1555  
QY 872 EPQSD--NDSSATCSADEVD-----GEPERQRMPPMDSKPSLLNPTGTSILVSSPLK 921  
DB 1556 DSEKDLLDSD---DDDIIELEECIISAMPTKS-----SRKAKKLAQTASKL-PPVPA 1604  
QY 922 PNPLDLP-----QLQHRAAVIP---PMVSC---TPCNIPIGTVPVSGYAL----- 959  
DB 1605 RKPQLPVYKLLPAQNLQAKHVSFTPGDDVPVYCVGEGTPINFSTATSLDLTIESPP 1664  
QY 960 -----YORHKAMHESALLEEQROREQIDLE 986  
DB 1665 NELATDGVRAGIOGSEFEKRDRTIPTEGRSTDDAQRGKISSIVTDLDDNKAEEGDLAE 1724  
QY 987 CRSSTSPCGTSKSNREWEVLQAPHQLITNLPEGVRLPTTRPTPPPPPLIPSSKTYVAS 1046  
DB 1725 CINSAMPKGSKHKPRVKKIMDVQQAASSTSGANKNOVDTKKKKTPSPVKPMPQNTYR 1784  
QY 1047 EKPSFIMGSGISOGTPTGYTLTSHNOASYTOETPKPSV-----GSGIS 1087  
DB 1785 TRVRKNTDSKYNVNTETP-----SDNKSCKPSLOTNAKAFNEKLPNNEDVRGTFA 1837  
QY 1088 LGLPRQOESAKSATLPYIKQEEFSPRQNSQOPEGLLVRQAHGVVVRGAPAGIQAEGSIT-R 1146  
DB 1838 LDSPHHTPIEGT--PYCFSRNDSLSLDDFDDVLSREKAELRKESKDSKAKVTCR 1895  
QY 1147 GTPTSKTSVES-----IPSLRGS---ITQGTALPQT-----GIPT-EALVKSGIS 1188  
DB 1896 PEPNSSQQAASKSOASIKHPANRAQSKPVLOKQPTFPQSSKDGPDGGAATDEKLONTAIE 1955  
QY 1189 RMPI-----EDSSPEKGREAAKSGHVI-----YEGSGHT-- 1219  
DB 1956 NTPVCFSRNSSLSSLDIDQNNNNKKESEPIKEAPANSQGEPSKQNSGVAPKSFHVED 2015  
QY 1220 -----LSYDN-----TKNAREGTRSPRTAHEISLKRSEYSEVGNIKOGM 1258  
DB 2016 TPVCFSRNSSLSSLDSEDDLLQECISSAMPKKKRPSRLKSEKQSPRVGGILAE DL 2075  
QY 1259 SMRESPPVASPLEGLICRALPRGSPHSDLKERTVLVSGTMOGTPTATTESFEDGLKYPKQI 1318  
DB 2076 TLDLKLQRPDSE---HAFSPGSENFQWKATQEGANSIVSSLHQAAAAA-----ACLSRQA 2128



```
Oy 1319 KRESPIRAFEAGAITKPKYDITTIKEMGRSIIHEIPRODILQESRKTPEVQVOSTPII 1378
Dy 2129 SSDSILSLKSGISLSPF-----HLTPDQBEKPTSNKGPRILKPG----- 2171
Oy 1379 EGSISQGTPIKFDNNSSQSAIKHNKSLITGP-----SKLSRGMPPLEIVPENIKVVERGK 1434
Db 2172 EKSTLEAKKIB-SENKGIKGGKVKYKSLITGKIRSNSEISSOMK--QPLPTNMPISISGR 2228
Oy 1435 YEDVKAGETVSRITSVSSGSPSVLRSLTLEAPKAQSLSPGIYDTSARTRPVSYQNTMSR 1494
Db 2229 TWIHIPGLRNSSTSPSVKGPPLKTPASKP-----SEGCAVTSRGTGPAKSEL-- 2282
Oy 1495 GSPMNRSTSDTIPPNKSTNRHKSTLTPTQRESIPAKSPV--PCVDPVVSHSPFDPHHR 1552
Db 2283 -SPITRQTSQIS-GSNKSSRSGSRDSTPSRPTQPLSRPMQSPGRNSI---SP-----GR 2333
Oy 1553 GSTAGEVYWSHLPTQDLPAMPFHALDPAAYLFQRLSTPTPGYQYQIYAMENRQOT 1612
Db 2334 NGISPPNKLSQLPRTSSPSTASTKSSGSKMSY-----TSPG-----RQLSQOQLTKQA 2382
Oy 1613 ILNDYITISQOQOVNLRPDVARGLSPREQPLGLPYPATRGIIDLTNMPPTILVPHPGTST 1672
Db 2383 SLKNASS-----IPRE-----SASGLNQMSN-----GNGSNKK 2413
Oy 1673 PPMDRITYIPGTQITFPPRPYNASMSPGHPHTHLAAASAERERE--REREKERERERIA 1730
Db 2414 VELSRM-----SSTKSSGSDSERPALVRQSTFIKEAPSLRRKLE 2457
Oy 1731 AASDLYLRPGS--EQGRPGSHGYVSPSPSVTQETMLQORSVFQGTNGTSTVTP-L 1787
Db 2458 ESASFESLSPSRDPSRTRQAQTPVLSPS-----LPDMSLSTHPSVQAG--GWRKLPNLL 2511
Oy 1788 DPTAQLRIMPLPAGPSTISQGLPASRYNTAADALAAALVDAASAPQMDVS-----KTKES 1842
Db 2512 SPTIE-----YNDGRPKRHDIA-----SHSESRLPINRAGTWKREHS 2552
Oy 1843 KHEAA--RLEENLR--GRSAVSEQQOLEQKTEKVRKSVQCLYTSTSAFFSGKQPQHSSV 1898
Db 2553 KHSSSLPRVSTWRRTGSSSILSASSESEKAKSEDERHVSMM-----2595
Oy 1899 VYSEAGKDKGPPKSRVEEELTRGKTTITANFIDVITIQIASDKDAREGQSQSDSS 1958
Db 2596 -----PAPROMKENOVPTKG-----TWKIKESDISPTGMSASOSAS 2631
Oy 1959 SSLSHRYETPSDAIEVISPASSAPPQEKLTQYQVVKVAKANQENPTROYEGPLHHYR 2018
Db 2632 SGAASGAESKPL-----IYQNAAPPVK-----TEDVWVRLEDPCPINPR 2670
Oy 2019 PQGESPSFQOQLPP-----SSQAEQMGQVPRTHRLITLADHI 2055
Db 2671 ---SGRSPGTGNTPPVIDSVSEKSSIKDKSKDTHGKQVSGSGSPVQTVGLETRLNSF 2727
Oy 2056 QCIITQDFARNQVSSQTPQOQPTST--FQNSPSALVSTPVTKTSNRYSPSQASQVHHQ 2113
Db 2728 VQV-----EAPQKGTBAKPGQSNPVSIAT--AETCAER--TPFSSSSSKHS 2773
Oy 2114 RP-----GSRVSPENLVKSRGSPKSPERSHVSSEPEYPTIS-PPQVPVVEHKQDQSLILL 2168
Db 2774 SPSCTVAARVTPFY-----NPSRKSSADSTARPQSIQITPVSTNTKKRDKTDI 2824
Oy 2169 SQRAEPAEQRNDARSPGCSISYL 2191
Db 2825 TESSGAQSPKRH-----SGSYL 2841

RESULT 11
ID ANK2_HUMAN STANDARD; PRT; 3924 AA.
AC Q01484; Q01485;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANKYRIN 2 (BRAIN ANKYRIN) (ANKYRIN B) (ANKYRIN, NONERYTHROID).
```

```
GN ANK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain stem;
RX MEDLINE=94075409; PubMed=8253844;
RA Chan W., Kordell E., Bennett V.;
RT "440-KD ankyrinB: structure of the major developmentally regulated
RT domain and selective localization in unmyelinated axons.";
RL J. Cell Biol. 123:1463-1473(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC TISSUE=Brain stem;
RX MEDLINE=91302466; PubMed=18330053;
RA Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT "Isolation and characterization of cDNAs encoding human brain
RT ankyrins reveal a family of alternatively spliced genes.";
RL J. Cell Biol. 114:241-253(1991).
RN [3]
RP REVISIONS.
RA Carpenter S.;
RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 463-495 FROM N.A.
RX MEDLINE=92009921; PubMed=1833308;
RA Tse W.T., Menninger J.C., Yang-feng T.L., Francke U., Sahr K.E.,
RA Lux S.E., Ward D.C., Forget B.G.;
RT "Isolation and chromosomal localization of a novel nonerythroid
RT ankyrin gene.";
RL Genomics 10:858-866(1991).
CC -1- FUNCTION: ANKYRINS ATTACH INTEGRAL MEMBRANE PROTEINS TO CYTO-
CC SKELETAL ELEMENTS; THEY BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN
CC BAND 4.2, TO NA-K ATPASE, TO THE LYMPHOCYTE MEMBRANE PROTEIN GP85,
CC AND TO THE CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND
CC DESMIN. ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO
CC THE CYTOSOLIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;
CC THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CC CELLS THROUGHOUT THE BRAIN.
CC -1- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
CC AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
CC AND FUNCTION (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 23 ANK REPEATS.
CC -1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z26634; CAB42644.1; -
DR EMBL; X58957; CAA40278.1; -
DR EMBL; X56958; CAA40279.2; -
DR EMBL; M37123; AAA62828.1; -
DR PIR; S14533; S14533.
DR PIR; A39643; A39643.
DR PIR; B39643; B39643.
DR PIR; S14569; S14569.
DR HSP; Q00420; IAWC.
DR MIM; 106410; -
DR InterPro; IPR000488; -
DR InterPro; IPR000906; -
DR InterPro; IPR002110; -
DR Pfam; PF00791; ZUS; 1.
DR Pfam; PF00023; ank; 22.
```





DR PROSITE; PS0176; ARM\_REPEAT; 1.  
KW Anti-oncogene; Phosphorylation; Coiled coil; Repeat.  
FT DOMAIN 1 728 LEU-RICH.  
FT DOMAIN 1 62 COILED COIL (POTENTIAL).  
FT REPEAT 125 260 COILED COIL (POTENTIAL).  
FT REPEAT 451 493 ARM 1.  
FT REPEAT 503 545 ARM 2.  
FT REPEAT 546 589 ARM 3.  
FT REPEAT 590 636 ARM 4.  
FT REPEAT 637 681 ARM 5.  
FT REPEAT 682 723 ARM 6.  
FT REPEAT 724 765 ARM 7.  
FT DOMAIN 739 2831 SER-RICH.  
FT DOMAIN 1130 1155 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 1864 1891 HIGHLY CHARGED.  
FT MUTAGEN 523 523 C-R: IN AN IQ-INDUCED COLON TUMOR.  
SQ SEQUENCE 2842 AA; 310530 MW; 3CB2EA8A34E9F47 CRC64;

Query Match 3.28; Score 406; DB 1; Length 2842;

Best Local Similarity 19.48; Pred. No. 5.7e-08; Indels 780; Gaps 116;  
Matches 461; Conservative 293; Mismatches 844;

Qy 297 NHARKQREKICQRYDQLMQAEWKKVDRINNPRKAKESKTREYIEKQFPPIKQREOQ 356  
Db 934 NFAKSESSNRTC-----MPYAKVEYKRSSNDLSNVTSSDGYGKRGOMKPSVESYEDD 988  
Qy 357 E-RFORVQORAGLSATARSEH-----EISEIDGLSEQENNEKQM 397  
Db 989 EGKFCYGYQPADLAHKTSHANHMDNGGELDTPIYSLKYSDEQLNSGRQSPSONERWA 1048  
Qy 398 RQLSVIPPMFDAQRRVKFTNMGLMEDPMKVYKDRQFMVMTDHEKEIFKDFIHPK 457  
Db 1049 RPKHVIDEIKQNGQORSQNTN-----FPVISEN-----TD-----DRHLFPQ 1089  
Qy 458 NFGLTASYLERKSPDCVLYLYLTKKNENYKALVRRNYKRRGRNQIARPSQEEKVEEK 517  
Db 1090 HFG-----QOECVSPY-----RSRGTSNETNRMGSSHAVNNVQSLQCE--DDY 1133  
Qy 518 EEDKAETK--EEEKDEBEKEDEKENTKEKIDGT-----AEETEREQAT 568  
Db 1134 BDRKPTNYSERYSEEEHEEERTNYSIKYN--EKKHVDQPIDYSLKYATYDSSQKPS 1192  
Qy 569 PRGKRTANSQRRKGRITRSMNTNAAAASA-----AAAAATEPPPPPLPPPEP 617  
Db 1193 FFSKTPSVQGTKEH--NSPSSSEASAPSNNAKRSQSLHPSSAQRNGQTPKGTACKVPS 1250  
Qy 618 ISTEPVET-----SRWT-----EEM-----EVAKGLV 641  
Db 1251 INQETMQTYCEDTPICFSRCSLSLSLSSAEDEIGCDQTTQEADSANTLQIAETKENDVT 1310  
Qy 642 EHRNWA-----AIKMWGTSKQACKNFYFNKRRHNDLNLQHQ-----KTSRKR 691  
Db 1311 RSAQDPASDPAVSOSTRTKPSRLOAGLASESARHKAVEFSFGAKSPSKSGAQTPKSP 1370  
Qy 692 EE-----RDVSOQESVASTYSAQDEDEIAEENEEPEDESEAVKPSDENATSRG 745  
Db 1371 EHYVQETPLVFSRCTSV--SLDSFESRSI--ASSVQSEPCSMVSGIVSPSLDP---SPG 1425  
Qy 746 NTEPAVELEPTETAPSTP-----SLAVPTK--PAEDESIVTQVNDISIAETAQM-- 796  
Db 1426 QTMP-----PSRSKTPPPPPPPVQVQTKREVFKTKVPAEQREGGPKQTAVSAVQRVQV 1480  
Qy 797 --DVDQOEHSAEE--GSVC-----DPPPATKADSDVDVVRV--PENHASKVEGD 839  
Db 1481 LPDADTLHLFATETPDGFCSSLSLSLSLDEPIQK-----DVELRIMPPVQ-----END 1531  
Qy 840 NTKERDLDRASEKVEPRDELVAQQAQINAOAPEQSD--NDSSATCSADEVD----- 890  
Db 1532 NGNETEPEQPEENENQDEK-----EKPDEKDLDDSD-----DDDIIEILECII 1578  
Qy 891 -GEFERQRMFFMDSKPSLLNTGTSILVSSPLKPNPLDLP-----QLQHRAAVIP 938

Db 1579 SAMPTKS-----SRKAKKLAQTASKL--PPVVARPSPQLPVYKLLPSQSRLOAQKHVSFT 1632  
Qy 939 -----PWVSC-----TPCNIPIGTPVSGYALQ-----RHIKAMHESALLEE----- 975  
Db 1633 GDDVPRVYCVEGTPIINFSTATLSLDLTIESPPNLAAGDGVRAVSQVGEFEKRDITIPTEG 1692  
Qy 976 -----OR-----QROEQIDL--ECRSSSTSPCGTSKSPNREWEVLQAPAPQL 1014  
Db 1693 RSTDEAQRGVSSIAIPDLGSKAEEDIIAECINSALPGRSHKPRVKKIMDQVOQAS 1752  
Qy 1015 ITNLPPEGVRLPTTRTPPPPLIPSSKTTVASEKPSFIMGSSISOGTPTGYLTLSHQASY 1074  
Db 1753 MTSSGNTKNQIDTKKKPTSPVKPMQNTVTRVRKNTDSKVNVTETTF--SDNKDS-- 1809  
Qy 1075 TOETPKPVSISISLGLPQOESAKSATLPYIKQEEFSPRSONSQ--PEGLLVRAHQEGVVR 1133  
Db 1810 -----KROSLKN-----NPKDLNDKLDP-----NEDVR 1833  
Qy 1134 GTAGIAQEGSITRTGTPTSKISVESIPSLRGSITOGTTPALPOTGIPTALVKGSIRMPIE 1193  
Db 1834 -----GFTFDSPHYAPIE-----GTPYCFSR-----NDSLSLDFD 1866  
Qy 1194 DSSPEKREEAASKGHVYIEKSGHILSYDNKNAECTRSPTAHEISLKRYSYVEGN 1253  
Db 1867 DDDVDLSREKAE-----LRKGE-----SKDS--EAKVTCHTEPSSSQSARKAQASTKHP 1915  
Qy 1254 IKQMS--HRESVPSAPLEGICRALPRGSPHSDLERTV-----LSGSTMQGTP-- 1301  
Db 1916 VNRGPSKPLLOEQP-----TFQSS--KDVDPGAATDEKLFONFAIENTPVCFS 1962  
Qy 1302 RATTESFEDGLKYPKQIKRESPPIRAFEGAITKGPYDGTITTIKEMGRSIEHPRQDILT 1361  
Db 1963 RNSLSLSDVDQENNEETGVPVDAEPANAQOP-----CK-----P 2001  
Qy 1362 QESKRTPEVQOSTRIIEGSSISOGTPIKFDNNSGQSAIKHN-----VKSITG----- 1409  
Db 2002 QASGYAPKSFH-----VEDTPVCFSRNSSLSDSDEDDLRECISSAMPKRR 2051  
Qy 1410 PSKLSR-----GMPLEIYVENIKVVERKYE-----DVKAGEIVRSRHT 1449  
Db 2052 PSRLKGGEWQSPRKVGSVLAEDLTOLKIQRPSEHGLSPDSENFDMKA----- 2102  
Qy 1450 SVSSGSPVLSRSTLHEAPKAQ--LSPGIYDDTSARRTPVSYQNTMTSRGSPMMNTSDVTI 1507  
Db 2103 --IQEGANSIVSSLHQAAAAACLSRQASSDSD--SLSLKGVSGLSGSP----- 2148  
Qy 1508 PPNKSTNHERKSTLTPTQRESIPAKSPVGVDPVVSHPDPHHRGSTAGEVYVSHLPTQ 1567  
Db 2149 -----HLTPDQEE-----KPF--TSHKGP----- 2165  
Qy 1568 LDPAMPFHRALDPAALVLFOROL--SPTPGYPSOYOLYAMENTRQTILNDYITSQ--M 1623  
Db 2166 -----RIKPEKSTLEAKKIESENKIGKGGKVKLSLITKIRSNSEISSQMQL 2217  
Qy 1624 QVNLRPDVARGLSPREOGLPYPATRGIIDLTNMPPTILVPHPGGTSTPPMDRITYIPG 1693  
Db 2218 QTNM--PSISRGRTW-----IHIPGVN-----SSSSTSPVSK-----KG 2250  
Qy 1684 TOITFPPRPYNASMSKSPCHPHTLAAAAAEREREREREKERERERERERERERERER 1743  
Db 2251 PPLKTP-----ASKSPSE-----GPVATTSPRGTKPAVKSLSPIITRQTS--HISGSKN 2297  
Qy 1744 QPGRPGSHGVRSPPSVRTQETMLQ-----RPSVFGTNGTSTVITPLDPTAQLRIMPL 1798  
Db 2298 GPSRSGS-----RDTSPRPTQOPLSRPMQSPGRNSISFGRNGIS--TP-NKLSQLPTSS 2350  
Qy 1799 PAGGPSISQGLPASRYNTAADALA-----ALVDAASAP-----QMDVSKTK 1840  
Db 2351 PSTASTKSSGSGKMSYTPSPGRQLSQNLKOTGLSKNASSIPRSESASKGLNNNGNS 2410  
Qy 1841 ESKHEAARLEENLRSNAVSEQQOQLEQKTLVEKRSVQCLYTSAPPSGKPOPHSSVYV 1900



Qy 309 QRYDOLMEAWKVDRIENPRKAK-----ESKREYVEKOPPEIRKOREQOERFORVGO 364  
Db 825 SASPLRQCIENGCVNAKTPANYKTSLETKSD-----TETESKVTSTVNR 874  
Qy 365 RGAGLSATIAKSEHEISIDLSQENNE-----KOMQOLSIVIPPMFADAEORRVKFI 418  
Db 875 SG---RSTEFNRNIQKL--VESKSEETVEICILKRGQKATLL-----QQR----- 918  
Qy 419 NNGLM---EDPMKYKDRQPMVWTDHEKIFDKKFTQHPKNGL-----TASLYERSKV 471  
Db 919 --EGEMKEIERPFETKE-----NIELKENDE--KMKAMKRSRTWQKCAPMSDLTDLKSL 970  
Qy 472 PDCVLYLYLTKNENYKALVRNNGKRGKRNQOITARPQSOEKEVEKEEDKAEK----- 524  
Db 971 PDTLM-----KOTARQNLILQTODHAKAPKSEKGIKTKPMQCSLQ 1011  
Qy 525 -----TEKKEEKKOBEKDEKED-----SKENTREKIDKIDTAET---E 562  
Db 1012 PEPINTPTHTKQOLKASLGKVGVEELLAVGKFTRTSGETHTHREPAGDCKSINTPKES 1071  
Qy 563 EREQATPRGRKTANSQGR-----KGRITRSTMTNEAAAAASAAA 601  
Db 1072 PKOILDPAARVTGMKKWPRTKPEEAQSLDLAGFKELFQTPGSEESMTDEKTTKIACKS 1131  
Qy 602 AATERPPPLPPPPPISTEPVETSRW-----TEEMEVAKKGLVEHGRNWAALAK 652  
Db 1132 -----PPESVDT--PTSTKQWPKRSRLKADVVEEFALURLKLTSPSAGK--AMLTPT 1177  
Qy 653 MVGKSEAOCKNFYNYKRRNLDNLQOHKOTSRRKPRE-----RVSQCESVASTVSA 708  
Db 1178 KPAGGDEKIDKAFMPTPVQKLDLAGTLPKSRQL--QTPKEKAQALEDLAGFKELFQTPG- 1235  
Qy 709 QEDIEDIASNEE-----ENPEDSEVAVKPSDESPENATSRGNETEPAVELEPTTETAPST 763  
Db 1236 --HTEELVAAGTKTIKPCOSQSDPDVDTSTKQKRSIRKADVEG---ELLACRNLMPSA 1292  
Qy 764 SPSLAVPTKPAEDSVETQVNDST-----SAETAQOM-----DV 798  
Db 1293 GKAMHTPKPSVGEEDKIIIFVGPVQKLDLTENLTGSKRRRQTPKEEAQALEDLTGFKEL 1352  
Qy 799 DQEHSAEAGSV-----CDPPATKADSVDEVVRPE----- 830  
Db 1353 FQTPGHTTEAANAAGKTKWPCSSPESADTPTSTRQPKTPLEKRDVQKELSALKILQ 1412  
Qy 831 -----NHASKYEGDNTKERDLDR--ASEKVEPRDEDLVAQOINAQRPESQDNDSSATC 883  
Db 1413 TSGETHTHDKVPGGEDKXSINAFRETAKOKLDP-----AASVTGSKRHPTKEKAQPL- 1464  
Qy 884 SADEDVGEPERQRMFMDSKPSLLNPTGSLVSSPLKPNLDLPLOH----- 932  
Db 1465 ---EDLAGMKELFQTPVCTDKPTTHEKTTKIACRS--QPDVDTPTSSKPSKRSRLKVD 1519  
Qy 933 -----RAAVTPMVVSCPTPCNIP--IGTPVSGYALYQ-----RHIAKMH 968  
Db 1520 VEEEFALRKRTSPSAGKAMHTPKPAVSVEKNYAFMGTPVQKLDLTENLTGSKRRLOTPK 1579  
Qy 969 ESAL-----LEQRQOEQIDLECRSS----- 990  
Db 1580 EKAQALEDLAGFKELFQTRGHTESEMTNDKTAACKSSQPDLDKNPASKRRLKTSIGK 1639  
Qy 991 -----TSPCGTSKSPNREVEVLQAPAHOLI-----TNLPEQVR 1023  
Db 1640 VGVKEELLAVGKLTQSGETHTHTEPGDGKS---MKAFESPQKILDSAAASLTGSKR 1695  
Qy 1024 LPTTRTPPPPLPSSTKTTVASKPSFIMGSGISQGTPTGYLTSNQAQSTQOE--TPKPS 1082  
Db 1696 QLRT-----PKGSEVPEDLAGFI---ELFQ--TP-----SHUKESMTNEKTKVS 1736  
Qy 1083 -----VGSISLGLPROQESAKSATLPIYIKOEFPSPSONSQPEGLLVRAOHEGVVRG 1134  
Db 1737 YRASOPDLVDTPTSKQPKPKRSRKAD-----TEEEFLAFKQTPSAG---KAMH--TDPK 1787

Qy 1135 TAGAIQEGSITRGTPTSKISVESIPSLRGSIQTGTPALPQTG--IPTALVKGSISSRMPIE 1193  
Db 1788 AVGEKDLINTLGTGPVQK-----LDQPCNLP-----GSRRL--- 1819  
Qy 1194 DSSPEKGREAAAGHVIYEGKSGHILSYDNKINAREGTRSPRTAHEISLRSKSYESVGN 1253  
Db 1820 ---QTRKEKAQ-----ALEELTGFRLEFQTPCTDNPTADEKTKKI--- 1857  
Qy 1254 IKQMSMRSPVSAPLEGLICRALPRGS---PHSDLKERTVLSGSIQGTFRATESEFD 1310  
Db 1858 -----LCKSPQSPAD-----TPTNTKQRPKRSKLKADVEEFL---AFRKLTPSAGK 1902  
Qy 1311 GLKYPKQIKRSPPIRAFEAGAIT-----KGKPYDGIITIKEMGR 1349  
Db 1903 AMHTPKAAVGEKDLINTFVGTPEVKLLDNLGSLKRRPQTPKEKAKALEDLAGFKELFQ 1962  
Qy 1350 S---IHEIPRODILITQESRKTPVQVQSTRIIEGSIQGTPIKFDNNSGQ-----SAIKHN 1402  
Db 1963 TPGHTEESMTDDKITEVCKSP-----QPDVKTPTSSKQRLKISLGRVG 2007  
Qy 1403 VKSLITGPSKLSRGMPLLEIIVPENIKVVERGKYEDVKAGETVRSRHTSVVSGPSVLRST 1462  
Db 2008 VKEEVLVPVKLUTQ-----TSGKTTQT--HRETAGDGKSI---KA 2041  
Qy 1463 LHEAPKAQLSPGIYDDTSAR--RTPVSYQNTMS--RGSPMMNRTSDVTIPPNKSTNHERK 1518  
Db 2042 FKESAKQMLDPANYGTGMRWPRTPKEEAQSLDLAGFKELFQTPD-----HTEE 2091  
Qy 1519 STLPTTORESIPAKSPVP--GVDPVVS--HSPFDPHHRGSTAGEV-----YWSHLP 1565  
Db 2092 ST--TDDKTTKIACKSPSPESMDTPTSTRRPKTPGLGRDIVEELSALKQLQTTHDKVP 2150  
Qy 1566 TQLOPAMPFHR---ALDPAAAAVLFQRLSPTPGYSQ--YOLYAMENRTQT--LLNDY 1617  
Db 2151 GDEKGINVRETAKQKLDPAASVTGSKRQ--PRTPKGAQPLEDLAGFKELFQTPVCTDK 2209  
Qy 1618 ITSQOMQVNLRPDYARGLSPREQPLGLPY-----PATRGI-----IDLNNPPTI-- 1662  
Db 2210 PTTHEKTKI---ACR--SQPDVGTPTIFKPSKRSRLKADVEEESALRKRTSPSVGK 2264  
Qy 1663 --LYPHPGGTSTPMDRITYIPGTQITFPPRPYNSASNSPGHPTHLAAAASAEERERER 1720  
Db 2265 AMDTPKPAGGDEKDMKFMGTPVQKLDLP-----GNLPG-----SKRWPTPK 2307  
Qy 1721 EKERERERIAAASDLYLRCSEQOPRPGSHGYV--RSPSPSVRTQETMLQOESVFOGT 1778  
Db 2308 EKAQALEDL--AGFKELFQTPGTDPTTDEKTTKACKSPQDPDVTPTASTQKPK----- 2361  
Qy 1779 NGTSVITPLDPTAQLRIM--PLPAGGPSISQGLPA-----SRYNATAADALAALVDAASAP 1832  
Db 2362 ---RNLKADVEEFLALRKRTSPSAGKAMDTPKPAVSDEKNINTFVETPVQKLDLGLNLP 2418  
Qy 1833 QMDYSKTKESKHEAARLEENLRS-----RSAVSEQOOLEOKTLEVERKRSVQCLYTSSAF 1887  
Db 2419 --GSKROPQTPKEKALEDLVGFKELFQTPGHTESMTDDKITEVCKS----- 2466  
Qy 1888 PSGKPOPHS--SVVTSEAGKDKGPPPKSYEE-----LRTCKTITITAAINFIDVITR 1939  
Db 2467 ---POPEFSTSSSKQRLKIPLVKVDKMEPLAVSKLRTSGETTQTHTE-----PTG 2517  
Qy 1940 QIASDKDARERGSSSSSSSLSSHR-----YETPSDAIEVIS 1977  
Db 2518 DSKSIKAFKESPKQILDPAASVTGSRRLRTRKEKARALEDLVDFKELFSAPGHTESMT 2577  
Qy 1978 -----PASSPAP-----POEKLOTYQPEVVKANQANDPTROYEGPL 2014  
Db 2578 IDKNTKIPCKSPPPPELTDATSTKRCPKTRKPKVEKLSAVERLTQTSQGSTHKEPA 2637  
Qy 2015 -----HHVRPOQESPSQOQLPPSSQAGMGVPRTHRLITLITLADHCQIIT 2060  
Db 2638 SGDEGIKVLKORAKKKNPVEEFSRRRPAPKEKAQPLEDLAGFKELTSETSGHTQESLT 2697  
Qy 2061 QDFARNQVSSQTPQOPPTST-----FQNSPSALYSTPVRTKTSNR--YSP 2104

Db 2698 ASKATKIPCEPPLVDDTTASTKRHLRTRVQVKEEPSAVKFTQTSGETTDADKEPA 2757  
Qy 2105 SQAQSVHQRGRSVLENLVDKSRGSRPGKSPERSHVSSEPPYPISPPOVPVVEHKODS 2164  
Db 2758 GEDGIRKALKESAKQTAPASVTVT-GSR--RRPRAPRESAQAIEDLAGFKDPAAGHTEES 2814  
Qy 2165 LLLLSQGAEPAEQRNDARSPGISYLPSPFTKLENTS-----PMVSKKOEIIFRKLNS 2218  
Db 2815 M-----TDDKTKIPCKSS-----PELEDATATSKRRPRTRAKQVEVKEELLA 2857  
Qy 2219 SG-----GGSDMAAQP-----GTEIFNLPA-----VTTSVSGSVSRGHSFADPASNGLLED 2265  
Db 2858 VGTUTQSGTTHTDKEPVGSGKGTAKQKAPAKRNVDAEDVIGSRQPRAPKEKAQPLED 2917  
Qy 2266 IIRKALMGSPDKVEDHGVNQPMGVVPGTAN-----TSVVTSGTTRREGDPPSPHSGGVC 2322  
Db 2918 -----LASFOE-----LSQTPGHTEELANGAUAFTSAPKQTPDPSGAPKLISRVL 2963  
Qy 2323 K-PKLISKNSRKSIPGQGYLGTERTP 2350  
Db 2964 RAPKVEPVGVVSTRDPVKQSQKNTSLP 2992

RESULT 14  
ID YCS3\_YEAST STANDARD; PRT; 1226 AA.  
AC P25357;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DE 15-DEC-1998 (Rel. 37, Last annotation update)  
DE HYPOTHETICAL 138.5 KDA PROTEIN IN RPS14A-GNS1 INTERGENIC REGION.  
GN YCR033W OR YCR33W OR YCR592.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92133166; PubMed=1776366;  
RA Wicksteed B.L., Roberts A.B., Sagliocco F.A., Brown A.J.P.;  
RT "The complete sequence of the unit YCR59, situated between CRY1 and  
MAT, reveals two long open reading frames, which cover 91% of the  
10.1 kb segment."  
RL Yeast 7:761-772(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91335897; PubMed=18772032;  
RA Jia Y., Slonimski P.P., Herbert C.J.;  
RT "The complete sequence of the unit YCR59, situated between CRY1 and  
MAT, reveals two long open reading frames, which cover 91% of the  
10.1 kb segment."  
RL Yeast 7:413-424(1991).  
CC -1- MISCELLANEOUS: THIS PROTEIN IS ENCODED BY A NON-ESSENTIAL GENE.  
CC -1- SIMILARITY: SOME, TO S.POMBE SPAC2E12.01.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
CC -----  
DR EMBL: X59075; CAA41799.1; -;  
DR EMBL: X59720; CAA42300.1; -;  
DR PIR: S15053; S15053.  
DR PIR: S19445; S19445.  
DR SGD: S0000629; YCR033W.  
DR InterPro: IPR001005; -;  
DR Pfam: PF00249; myb-DNA-binding; 2.  
DR PROSITE: PS50090; MYB\_3; 1.  
KW Hypothetical protein.  
FT DOMAIN 594 599 POLY-SER.

FT VARIANT 305 305 T -> A.  
FT VARIANT 375 375 H -> S.  
FT VARIANT 404 404 H -> O.  
FT VARIANT 435 435 A -> V.  
FT VARIANT 442 442 D -> N.  
FT VARIANT 644 644 Y -> H.  
FT VARIANT 801 801 E -> D.  
FT VARIANT 806 806 K -> Q.  
FT VARIANT 831 831 I -> V.  
FT VARIANT 846 846 E -> G.  
FT VARIANT 851 851 A -> D.  
FT VARIANT 855 857 GVR -> AVQ.  
SQ SEQUENCE 1226 AA; 138503 MW; 8D133A0918658E53 CRC64;

Query Match 3.1%; Score 390; DB 1; Length 1226;  
Best Local Similarity 20.2%; Pred. No. 8.6e-08;  
Matches 222; Conservative 181; Mismatches 368; Indels 326; Gaps 52;

Qy 31 NTRHQQFAVPDYRSSHLEVSQAOLQQOQQOQLRRRPSL-LSEFHFGS--DRPQ---- 83  
Db 247 NSIHQRE--PFWKAN-----STILKSTHSQS---SPSLHTKKFHDANKLDREASVK 294  
Qy 84 -----ERRT-SYE-----PFHPG----- 95  
Db 295 VETPSKDETKTISYHDNNFPKRKSVSRKNAPLEPDNIKVGEDALGKKEVHKGSGREIAKE 354  
Qy 96 -PSVP---DHDSLESKRPRLEQVSDSHFQVSAAVLPLVHPLPEGLRASADAKDPFGG 151  
Db 355 HPTVPVKMKEHDELEARAKVVKNI-----NIDGKQDEIWT 389  
Qy 152 KHEAPSPISGQPCDDQNASPSKLSBELIOSMDRVDV-EIAKVEQ---OILKLKKKO 206  
Db 390 AKTVASA-----VEVSKEHKELTRSVKESPEIRYERAYDPKALKTDAIK 437  
Qy 207 QOLEEEAAKPPPEP-----EKVPSPPPVEQKHRSIVQIYDENRKK--ABEAHKEIFGLGPK 260  
Db 438 LTVDDDNKSYEEPLEKVEGCIFFPLP---KAETRLWELKNQKRNKIISKQYLLKKAIRNF 494  
Qy 261 VELPLYNQPSDTKYVHENIKTNQVWRKKLILFFRRNRHARKOREQKICQORDOLMEANEK 320  
Db 495 SEYFFYQAQ---NKLIIHQ--QATGLILTKIISKIRKEHLKK---INLKHDFDLOKKYK 546  
Qy 321 KVD---RIENPNRRKAKESKTREYVEKQFPEIRKOREQOERQFQVQORGA--GLSATIAR 375  
Db 547 ECEILTILKLSNLRKEEITENKKEH-----ELMEOKRREEGIEETEKEKSLRHPSSSSSR 600  
Qy 376 SEHEISEIIDLGSQEN-----NEKOMQLSVIPPMFMDAQR-RVKFINMGLMED- 426  
Db 601 RRNRADFVDD--AEMENVLLQIDPNKYHYQAAATPPLLDPIRKYSYKFCDDVNNLVTDK 658  
Qy 427 ---PMKYVKDRQFMNVVTDHKEIFKDKFIOHPKNFGLIASYL-ERKSVPCVLYYYLTK 482  
Db 659 KLWASRIILKAS--DNFTDHEHSILFLEGYLTHPKKFKGISHYMGGLRSPEECVLHYRTK 716  
Qy 483 KNEYKALVRRNYKGRGRNQOIAIRPSOEKVEEKEEDKAEKTEKKEKK--DEEEXDE 540  
Db 717 KTVYKQLLIDKNKRRK-----MSAAAKRRKRKERSNDEEVEVDESKESTWIDKEESE 772  
Qy 541 KEDSKEN-----TKEKDIDGTAEETEEREQATPRGRKTAN--SQGRKRGRITRSM 590  
Db 773 -NNAEENQPVLVQGVSEVKGDPGLGTPEKVE--NMIEKRGEFAGELENAERVNDLKAHD 829  
Qy 591 NEAAAAASAAAAATEEPPLPPP-----PEPI-----STEPVET 625  
Db 830 EIGESKSSVIETNNEVQIMAPKGVGRNGYVPEETKELDFSLLENALQKKHKAPEHKT 889  
Qy 626 SRWTEEMEYAKGLVEHGRNWAIAAKMVGTKSEAQCQKFNFKRRNRNLNLLQOHHQK 685  
Db 890 SYWSVRESQVLPPELLKEFGSOWSLSEKLGHKSTWYRNRYQORNAARNGWKLIVDETDLK 949  
Qy 686 TSRKPREEDRSDQCESVASTVSAQEDIEDIASNEEENPDESEVAVK--PSEDSPE---- 739

Db 950 -----RDGTSSES-----QQSQILQPERPNINAYSNIIPQORPALGYF 989  
Qy 740 -----NATSGNTEPAVE-----LEPTETAPSTSPSLAVPSTKPAED- 777  
Db 990 VQOPHTGHNTSISSIDGSRFPGDFHDTSKISAPLTTLPPLPSIQPPRSEMAEPT 1049  
Qy 778 -----ESVETQVNDISIAETAQMDVDQOQHSAGEGVCDDPPATKADSDVDVEVR 828  
Db 1050 VTDLRNRLDHDIT- LADAASSVT-----NNQNFERNAI----- 1084  
Qy 829 PENHASKVEGNTKVERDLDRASEKVEPRDEDLVVAQINAQORPEPQDNDSSATCSADE- 887  
Db 1085 -----DIGRKSSTIS-----NLLNDSRSMKSFQASRHEA 1116  
Qy 888 DVDGEPRQRMFMDKSPKLLNPGTSILVSSPLKP-----NPLDLPOL--QHRAVAP 938  
Db 1117 QLEDTSPNNIVQEKRNITTPRSS-SISALLAPVNGQSNPDGRPLLPFHAIQSGT 1175  
Qy 939 PMVSCTCPNIPIGTPVS 955  
Db 1176 PTF---PLPAPRTSPIS 1189  
RESULT 15  
PGCV\_CHICK  
AC Q90953; Q90945; STANDARD; PRT; 3562 AA.  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE (CHONDROITIN SULFATE PROTEOGLYCAN PRECURSOR (LARGE FIBROBLAST PROTEOGLYCAN)  
GN CP5G2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN-WHITE LEGHORN; TISSUE=Limb bud;  
RX MEDLINE=93300846; PubMed=8314802;  
RA Shinomura T., Nishida Y., Ito K., Kimata K.;  
RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan  
RT expressed during chondrogenesis in chick limb buds. Alternative  
RT spliced multiforms of PG-M and their relationships to versican.";  
RL J. Biol. Chem. 268:14461-14469(1993).  
CC -!- FUNCTION: MAY PLAY A ROLE IN INTERCELLULAR SIGNALING AND IN  
CC CONNECTING CELLS WITH THE EXTRACELLULAR MATRIX. MAY TAKE PART IN  
CC THE REGULATION OF CELL MOTILITY, GROWTH AND DIFFERENTIATION. BINDS  
CC HYALURONIC ACID.  
CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; V0 (SHOWN HERE) AND  
CC V1; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: PRECHONDROGENIC CONDENSATION AREA OF  
CC DEVELOPING LIMB BUDS.  
CC -!- DEVELOPMENTAL STAGE: DISAPPEARS AFTER THE CARTILAGE DEVELOPMENT  
CC (BY SIMILARITY).  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC use European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X60226; CAA42787.1; -

DR EMBL; D13542; BAA02742.1; -.  
DR HSP: P00740; IIXA.  
DR InterPro; IPR000152; -.  
DR InterPro; IPR000436; -.  
DR InterPro; IPR000538; -.  
DR InterPro; IPR000561; -.  
DR InterPro; IPR001304; -.  
DR InterPro; IPR001881; -.  
DR InterPro; IPR003006; -.  
DR Pfam; PF00008; EGF\_2.  
DR Pfam; PF00193; Xlink; 2.  
DR Pfam; PF00047; ig; 1.  
DR Pfam; PF00059; lectin\_c; 1.  
DR Pfam; PF00084; sushi; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 2.  
DR PROSITE; PS01241; LINK; 2.  
DR PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
DR PROSITE; PS50041; C-TYPE LECTIN\_2; 1.  
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;  
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
KW Hyaluronic acid; Alternative splicing.  
FT SIGNAL 1 26 POTENTIAL.  
FT CHAIN 27 3562 VERSICAN CORE PROTEIN.  
FT DOMAIN 37 136 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 166 243 LINK 1.  
FT DOMAIN 264 345 LINK 2.  
FT DOMAIN 3254 3290 EGF-LIKE 1.  
FT DOMAIN 3292 3328 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 3330 3459 C-TYPE LECTIN.  
FT DOMAIN 3460 3518 SUSHI.  
FT DISULFID 44 129 BY SIMILARITY.  
FT DISULFID 171 242 BY SIMILARITY.  
FT DISULFID 195 216 BY SIMILARITY.  
FT DISULFID 269 344 BY SIMILARITY.  
FT DISULFID 293 314 BY SIMILARITY.  
FT DISULFID 3258 3269 BY SIMILARITY.  
FT DISULFID 3263 3278 BY SIMILARITY.  
FT DISULFID 3280 3289 BY SIMILARITY.  
FT DISULFID 3296 3307 BY SIMILARITY.  
FT DISULFID 3301 3316 BY SIMILARITY.  
FT DISULFID 3318 3327 BY SIMILARITY.  
FT DISULFID 3334 3345 BY SIMILARITY.  
FT DISULFID 3362 3454 BY SIMILARITY.  
FT DISULFID 3430 3446 BY SIMILARITY.  
FT DISULFID 3461 3504 BY SIMILARITY.  
FT DISULFID 3490 3517 BY SIMILARITY.  
FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 235 235 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 529 529 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 709 709 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 948 948 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1409 1409 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1479 1479 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1523 1523 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1530 1530 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1625 1625 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1988 1988 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2088 2088 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2089 2089 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2507 2507 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2642 2642 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2748 2748 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2762 2762 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 3069 3069 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 3194 3194 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 3232 3232 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 3545 3545 N-LINKED (GLCNAC. .) (POTENTIAL).



FT. VARSPLIC 485 1411 MISSING (IN ISOFORM V1).  
SQ. SEQUENCE 3562 AA; 388078 MW; 9BC566E88C1602D2 CRC64;

Query Match 3.08; Score 374.5; DB 1; Length 3562;  
Best Local Similarity 18.18; Pred. No. 1.1e-06;  
Matches 511; Conservative 379; Mismatches 1102; Indels 833; Gaps 118;

QY 9 NOGAFSTEQ-SRYPHSHVQ-----YFPNTRHOQ--EFAVPDVRSSHLEVSQASQL 56  
Db 781 SEGRFGSEKETHPPVSGMLQTDKQVYETEESHKRIELDTEDDISMEPTSSPGQI 840  
QY 57 LQ-----OQOQOLRRPRLSSEFHGSDRPOERRTSYFFHPGPS----- 97  
Db 841 IEYTKHLGAPSAVTDKTSMTAETESDEEVVSADFDQTKGT-BVFHTSSSLDLEKF 899  
QY 98 -----PVDHDS-----LESKRRLQVSDSHFQORVSAAVLPLVHPLPEGLRA 139  
Db 900 TLSKIPEDSSATVKSPSSSGTVLPTAVATVLEVTHEADETSGYVLNMFTSTPEGEOR 959  
QY 140 SADAQKDPATFGGKHEAPSSPISGQDQNASPKLSKEELIQSDMDRVDREIAKVQOI 199  
Db 960 KA-TEKSPA-----TSADEVSTGTEISKYTMTEG-QGISVTSAAEKESV 1002  
QY 200 LKLLKKQOQ-----LEEAAKPEPEKPVSPPPVE-----QKH 232  
Db 1003 AALQEREQPSVGLPETKEPKFTDVTETETVTPQREGDSLVPVTVGSEDIGEMQVTDH 1062  
QY 233 RSIVQIYIDENRKAABEAHKIFELGPKVLPVLYNQSDTKVYHEN---TKNQVMRKL 289  
Db 1063 TSFDSIIHTEATVTSKASEVF---PK-ELSTKDQDRELGTAMGSTLPVTSVQMHEQKT 1117  
QY 290 ILFPKRRNHARKOIKORICYDOLMAEWKQVDRI---ENNPRAKAKESKTREYVEKQF 346  
Db 1118 TAGFESQPTTQEKHDEMGDAYDENYPTATLSVPALMLTEYGQVSGVPEVTSRSLHTGT 1177  
QY 347 PEIKOREQOQRFQVQORGAGLSATIARSEHSEIETIDGLSEQENNEKOMQLSVIPPM 406  
Db 1178 PKAETATDQEKITEAVPVTFGTQAKVYESK-----GTTTREEDRDVGSNWSLPPH 1229  
QY 407 MFDAEQRRVFINNGLMEDPMKVYKDRQPNVWTDHEKEIFDKDFIQHPKNGLIASYL 466  
Db 1230 TMLSSPSTAGSISLLTLGASPS-----OTPEGGISBELE 1264  
QY 467 ERKSVPCVLYLYLTKNENYKALVRNRYGKRRGNQOIAARPQOEKVEEKEKAEKTE 526  
Db 1265 EVKIVP-----FSSRATDKTYI-----SDLTSSISAVDKIQTSASKPVSS 1308  
QY 527 KKEEKKDEEKEKED-----SKENTKEKDKIDGTAETEEREQATPRGRKKTANSQ 578  
Db 1309 KSPRIIPEDEEVTSSDIIVIDESISPSKASAEDDLTKMKVEPE----- 1352  
QY 579 GRRKGRITRSMTEAAAAAATAATEEPPLPPPPPEPTEPVTSRWTEBEMEYAKK 638  
Db 1353 -----IDKEYFTSSATAVARTAPTPTVNEATEALQOPEVSPSTSHPOSGTDIR-- 1400  
QY 639 GLVEHGRNAAIAKMGVTKSEACKNFYFKRRHNDLNLQOHKQTSRKPREERDVSQ 698  
Db 1401 -----LVYLIQITGNDTDHPVNEPLDLFSRH-----ILPHAVDETHDAESAQTEP 1445  
QY 699 CESVASTVSA-----QEDIEDIE-----ASN 718  
Db 1446 CTSDSVDSSEYIILDPPFNFMDFEEDDECENTDVTTPPALQFNGKQVTSAPKST 1505  
QY 719 EENNPEDSEYEAVKPSDSEPNATSRGENTPAVELEPTTETAPS-----TSPSL 767  
Db 1506 KAEARSQIESVAHSKNVTFQINETNTIIEETASGTMQPSKAGEVGAEVTOPTA 1565  
QY 768 AVPSTKPAEDSEVETQVNDISAFETA---EQMDVDQOE---HSAEAGSVCDPPPPKADKS 821  
Db 1566 DVAMLEPVYSESEVTTDKYLEITSVYEQSPKKNKETVMVHGHTEESSTKD---TK--- 1618  
QY 822 VDVEVRVPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQOINQAORPEPOSDNDSSA 881

Db 1619 ---NLLLTITNESS---GDGSTESDLRS-----VFTEI-----L 1646  
QY 882 TCSADEVDGEPEPQRMFPMDSKPSLLNPGTSGILVSSPLKPNP-----LDLPOLQHRANV 936  
Db 1647 TMSHED-----SEKISHTSTVPTILSVERSAVTAAPSADSDTATGIDVKDL----- 1694  
QY 937 IPPMVSTPCNIPIGTIPVSGVALYORHIKAMHESALLLEEQRQOEQI----- 983  
Db 1695 IPKCGTATPGN-----YKSTIKLDAEFPFSPSNPEATSHHTKPDMTASSFVILE 1743  
QY 984 ---DLECRSTSPGTSKSPNREWEVLQAPAPHOLITNLPEGVRLPTTRTPRPPPLIPSS 1040  
Db 1744 GSGDVEENSTLASAMTTETAETLSVQD-----TSLGSGTVLPTETIS-----VTIS 1790  
QY 1041 KTVIVASKEPFIIMGSGISQGTPTGYILSHNQASVYQETPKPSVGSISLGLPQOESAKSA 1100  
Db 1791 EITPALPGGTRILYSTFDQSEATVST-----NFVSELIMEQVVGVSVATEKKEVEDEK 1845  
QY 1101 -TLPYIKOE-----BESP-----RSONSOPEGL-----LVRAQHEGVVGTGA 1136  
Db 1846 QTVVYSQEIISTTDKAKKSELDEFGSTTNEVRTVTSQEBTPPLREIVPTIGTMHSEIKKVTGA 1905  
QY 1137 GA-----IQEGS-----ITRGTSKISVESIP---SLRGTSITQGTALPOTGIPT 1180  
Db 1906 TPFLREKLFINEGSAEAPADLFAGSTPKRVVSTDSPTDSSGSDIDVITESATLTSVPSR 1965  
QY 1181 ALV-----KGSISRMPIEDSSPEKGBREBAASKHVIYEGSKSHILSYDNKKNAREGTR 1233  
Db 1966 SVIETQTVKHEGNINVISLKNTTTEYEE-----HI---GTGGPVTSVSS-----TGSD 2012  
QY 1234 SPRTAHSILKRSVESYEGNIKQKMSRMSPVSPAPLEGLICRALPRGSPHSDLKERTVLS 1293  
Db 2013 GLTEESEVAITEMS-ENVFSTENQGEPTQE-----AVPTTAPSDIASKRL--- 2055  
QY 1294 GSIMQGTPTATTESFEDGLKYPKOIKRESPPIRAFEGAITKGKPYDGTITTIKEMGRSIEH 1353  
Db 2056 GSRREVTSHT-----PVIRT----- 2071  
QY 1354 IPRQDILITQSRKTPVYVQSTRPIIEGSIQGTPIKFDNNSGSAIKHNKSLITGSPKL 1413  
Db 2072 ---KOLETAETVTSPPESV-----VNNSLDTMVTHTGTIRAVAESTES 2110  
QY 1414 SRGMPPLEIYP-ENIKVVERGKYEDVAKAGETVRSRHSVSVSSGPS---VLRS----- 1461  
Db 2111 KKGKGSFSAVSLGLKLMIEHSGEELK---VDSSTTKLMSNGPTEKLCSHFSFFDOGS 2166  
QY 1462 ---TLHEA-PKAQLSPGIYDDTSARPTPVSY-QNTWSRSGSPMMNRTSDVTIPPNK---S 1512  
Db 2167 GEATLETSETKASVSP-----TGKPEPQEQYGRKTVSMPSAVVHA---YTAEPNELVTS 2218  
QY 1513 TNHERKS--TLTPTQRESIPAKSPVPGVDPVWVSHSPF--DPHHRGSTAGEVYWSHLPTOL 1568  
Db 2219 TEHDITSQVITDTEMEKKAANELT--VTSFATNLPLSEDVHS-----WEDRPREI 2267  
QY 1569 DPAMPFFRALDPAAAA---YLFQRLS-----PTFGPSPQYQIYAMENTROTIL- 1614  
Db 2268 LP-----KAIESSEATEDPFFISTQANHEHVEFLSVPTIRPHSEENKVEAESDEKILLP 2322  
QY 1615 --NDYIT-----SQMQVNLRP--DVAR-GLSPREQLGLPYPATR 1650  
Db 2323 FNNDRVTESAVIERKYLSSPFTDTEQEELVQNIFFTEDIPRLFTPKEE----- 2372  
QY 1651 GIIDLTNMPTT--ILVPHPGTSTPPMDRTIYIPG-----TQITFPPRPYNSASMS 1700  
Db 2373 -----KPTNNELISDPLFSGQSGDEFTVIPSVESLAVKETTNTLSPPFPFASVGP 2424  
QY 1701 ---GHPTHLAAAASAEARE--REREKERERERIAAASDLYLRPGSQBPGRPGSHGVRS 1756  
Db 2425 KLSDTKQVFESGSDTSDNAEINEEITTTAAELTETATYSMATSSPALSEESSHSNSDKD 2484  
QY 1757 PPSVTRQETMLQORPVSFOGTNGTSTVITPLDPTAQLRIMPLPAGGPSISOGLPASRYNT 1816

Db 2485 ITHYFLVIEDPYNKEMDHRHNGENTS-----RPLTPG---DVSLESSHML 2528  
Qy 1817 AADALAALVDAASAPOMDVSK---TKESKHEAARLEENLRSA--AVSEQ-----Q 1864  
Db 2529 TTDDVTPVSVILSTPYLEMKGLSATATKMPRSRVLPSSGEGSGWGVDSFAPDILTH 2588  
Qy 1865 QLEQKTLVEKRVSQCLYTTSFAFSGPKPQPHSSVVSAGKDKGPPPKRSRYEEELTRGK 1924  
Db 2589 STAFSVMEVE-----LTASSHIFG-----VSEVMTTHVPGDGS-----Q 2623  
Qy 1925 TTIT--AANFD--VIITRQIASDKARERGSSQSSSSLS-----1962  
Db 2624 TVITGLASLFTTEERIEVANRTAADPKTGTSEELTSDTGMSLDIIPVDDRRHVTLNVSU 2683  
Qy 1963 -----SHRYTPSDATEVISPASSAPPOE--KIQYQPVVVKANQAEADPTR---QY 2010  
Db 2684 GDYLIERLQIPSEKTIIDMHSKMPEDIIISVQTPNPNLVIRSTQVSDDNMKAEDKY 2743  
Qy 2011 EGPLHRYRPOQES-----PSPQOQLPSSQAEGMGQVPRTHRLITLADHICQIITQDFA 2064  
Db 2744 DSIINFSTVEENSGDNLSTTSIQPSSSESVTAGHGPKLVKDLGSGYAMQFATETLT 2803  
Qy 2065 R-----NOVSSOTPOOPTSTFONSPSALVSTPVTKTNSRNPSPESQASQV 2110  
Db 2804 TTVLNEGLIFLTPVPSLVSPHMPHESKESEFEAKHIGRTST-----TDDVVEPYTSANN- 2857  
Qy 2111 HHQPGSRVSNENLVDKSR---GSRPKCKPSERSHVSSEYEPISPPQVPPVHEKQ---- 2162  
Db 2858 -----QVITDQKTHMSIGFGMGQEEG-DKKPMIPSLTDLTNETEKALTTD 2905  
Qy 2163 --DSLILLISQGAEPAE--QRNDARSPGSIYL--PSFETKLENTSPMVKSKQBFIRKLN 2917  
Db 2906 TFDVSMVTQMSQSHATVSSSSSEKSHVTVMQTKASTEYEEDSV-----SLN 2955  
Qy 2218 SSGGSDMAAQAQCGEIFNLPATVTSVSGSRGHSFADPNASNLGLEDIIRKALMGSDFD 2277  
Db 2956 SVSQNKSSVTWLVNGVSKYPIIIPSTSSAKDSQSDHSSD-----GTFKE 3003  
Qy 2278 KVEDGVVMSOPMGVPTGANTSVVTSGETRR--EGDPSPHSGVCKPKLISKNSRKS 2335  
Db 3004 VSSDMAATYKPTDLDITVTSLLVFPPESEISISTESTPH-----FNKFTVTERS 3054  
Qy 2336 KSPIPGOGYLGTERPSSV-----SVHSEGDYHROTPGAWEDRPS-----STGSTOFF- 2384  
Db 3055 EETESSVNDLIIENATVSGDSPSIH--DY--PTAFWNGERTSTDPVKLSTIEVEFSS 3109  
Qy 2385 ---YNP-----LTMRLMSLTPPTPIACAPSAVNQA--PHQONRIWEREPAPLLSAQE 2433  
Db 3110 ERVKNPQESDRSTERERPRLSAPVSDSPNSIEVGFKPDQ-----EAVTMLTSSLE 3162  
Qy 2434 TLDSD 2438  
Db 3163 PLDRS 3167

## RESULT 16

ID\_HUMAN  
AC P25054; Q15162; Q15163; PRT; 2843 AA.  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ADENOMATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN).  
GN APC OR DP2.5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91335210; PubMed=1651563;  
RA Kinzler K.W., Nilbert M.C., Su L.-K., Vogelstein B., Bryan T.M.,  
RA Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McKechnie D.,  
Finnear R., Markham A., Groffen J., Boquski M.S., Altschul S.F.,  
Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.;  
"Identification of FAP locus genes from chromosome 5q21."; Science 253:661-665(1991).  
[2]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=Fetal brain;  
RX MEDLINE=91330307; PubMed=1678319;  
RA Joslyn G., Carlson M., Thliveris A., Albertsen H., Gelbert L.,  
Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,  
Sargeant L., Krapcho K., Wolff E., Burt R., Hughes J.P.,  
Warrington J., McPherson J.D., Wasmuth J., Le Paslier D.,  
Abderrahim H., Cohen D., Leppert M., White R.;  
"Identification of deletion mutations and three new genes at the  
familial polyposis locus."; Cell 66:601-613(1991).  
[3]  
RP ASSOCIATION WITH CATENINS.  
RX MEDLINE=94082295; PubMed=8259519;  
RA Su L.-K., Vogelstein B., Kinzler K.W.;  
"Association of the APC tumor suppressor protein with catenins."; Science 262:1734-1737(1993).  
[4]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=94154728; PubMed=8111410;  
RA Nagase H., Nakamura Y.;  
"Mutations of the APC (adenomatous polyposis coli) gene."; Hum. Mutat. 2:425-434(1993).  
[5]  
RP VARIANTS FAP.  
RX MEDLINE=91335211; PubMed=1651563;  
RA Nishisho I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,  
Koyana K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,  
Petersen G., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,  
Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.;  
"Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients."; Science 253:665-669(1991).  
[6]  
RP VARIANTS FAP.  
RX MEDLINE=93285030; PubMed=1338904;  
RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,  
Miki Y., Mori T., Nakamura Y.;  
"Somatic mutations of the APC gene in colorectal tumors: mutation  
cluster region in the APC gene."; Hum. Mol. Genet. 1:229-233(1992).  
[7]  
RP VARIANTS FAP.  
RX MEDLINE=93244793; PubMed=1338691;  
RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,  
Nakamura Y., Horii A.;  
"Somatic mutation of the APC gene in gastric cancer: frequent  
mutations in very well differentiated adenocarcinoma and signet-ring  
cell carcinoma."; Hum. Mol. Genet. 1:559-563(1992).  
[8]  
RP VARIANT FAP W-1348, AND VARIANTS D-1118; M-1292; V-1304 & S-2502.  
RX MEDLINE=93250848; PubMed=1338764;  
RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,  
Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,  
Baba S., Nakamura Y.;  
"Screening for germ-line mutations in familial adenomatous polyposis  
patients: 61 new patients and a summary of 150 unrelated patients."; Hum. Mutat. 1:467-473(1992).  
[9]  
RP VARIANT FAP TRP-99.  
RC TISSUE=Peripheral blood lymphocytes;  
RX MEDLINE=95134544; PubMed=7833149;  
RA Dobbie Z., Spycher M., Huerliman R., Ammann R., Roth J.,  
Mueller A., Mueller H., Scott R.J.;  
"Mutational analysis of the first 14 exons of the adenomatous  
polyposis coli (APC) gene."; Eur. J. Cancer 30A:1709-1713(1994).

[10] VARIANT FAP GLY-722.  
 RP MEDLINE-95135430; PubMed=7833931;  
 RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,  
 RA Romio L., Pilia S., Prete F., Marenzi C., Guanti G.;  
 RT "Four novel mutations of the APC (adenomatous polyposis coli) gene in  
 RL FAP patients.";  
 RL Hum. Mol. Genet. 3:1687-1688(1994).  
 [11]  
 RP VARIANT FAP ILE-171.  
 RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,  
 RA Romio L., Pilia S., Prete F., Marenzi C., Guanti G.;  
 RL Hum. Mol. Genet. 3:1918-1918(1994).  
 [12]  
 RP VARIANT FAP ILE-171.  
 RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,  
 RA Romio L., Pilia S., Prete F., Marenzi C., Guanti G.;  
 RL Hum. Mol. Genet. 3:1918-1918(1994).  
 [13]  
 RP VARIANT LYS-1307.  
 RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,  
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,  
 RA Antin-Ozerkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,  
 RA Gallinger S., Kaback M., King M.C., Woodage F., Brody L.C.,  
 RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;  
 RT "The APC I1307K allele and breast cancer risk.";  
 RL Nat. Genet. 20:13-14(1998).  
 [14]  
 RP VARIANT LYS-1307 AND GLN-1317.  
 PC TISSUE-Peripheral blood;  
 RX MEDLINE-98400248; PubMed=9731522;  
 RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,  
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,  
 RA Antin-Ozerkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,  
 RA Gallinger S., Kaback M., King M.C., Woodage F., Brody L.C.,  
 RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;  
 RT "The APC I1307K allele and breast cancer risk.";  
 RL Nat. Genet. 20:13-14(1998).  
 [15]  
 RP VARIANT LYS-1307.  
 RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,  
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,  
 RA Antin-Ozerkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,  
 RA Gallinger S., Kaback M., King M.C., Woodage F., Brody L.C.,  
 RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;  
 RT "The APC I1307K allele and breast cancer risk.";  
 RL Nat. Genet. 20:13-14(1998).  
 [16]  
 RP VARIANT LYS-1307.  
 RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,  
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,  
 RA Antin-Ozerkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,  
 RA Gallinger S., Kaback M., King M.C., Woodage F., Brody L.C.,  
 RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;  
 RT "The APC I1307K allele and breast cancer risk.";  
 RL Nat. Genet. 20:13-14(1998).  
 [17]  
 RP VARIANT LYS-1307.  
 RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,  
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,  
 RA Antin-Ozerkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,  
 RA Gallinger S., Kaback M., King M.C., Woodage F., Brody L.C.,  
 RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;  
 RT "The APC I1307K allele and breast cancer risk.";  
 RL Nat. Genet. 20:13-14(1998).  
 [18]  
 RP VARIANT FAP PRO-1184.  
 RA Lamlum H., Ilyas M., Rowan A., Clark S., Johnson V., Bell J.A.,  
 RA Frayling I.M., Efstathiou J., Pack K., Payne S., Roylance R.,  
 RA Gorman P., Sheer D., Neale K., Phillips R., Talbot I.C., Bodmer W.F.,  
 RA Tomlinson I.P.M.;

RT "The type of somatic mutation at APC in familial adenomatous polyposis  
 is determined by the site of the germline mutation: a new facet to  
 Knudson's 'two-hit' hypothesis.";  
 RL Nat. Med. 5:1071-1075(1999).  
 CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-  
 CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE  
 CC ALLOWS THE DOWNGULATION OF CYTOPLASMIC BETA-CATENIN.  
 CC -1- SUBUNIT: FORMS HOMOLOGOMERS AND ASSOCIATES WITH CATENINS.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES TYPES.  
 CC -1- PTM: PHOSPHORYLATED BY GSK-3B.  
 CC -1- DISEASE: DEFECTS IN APC ARE A CAUSE OF FAMILIAL ADENOMATOUS  
 CC POLYPOSIS (FAP) AND GARDNERS SYNDROME (GS), THAT CONTRIBUTE TO  
 CC TUMOR DEVELOPMENT IN PATIENTS WITH NONINHERITED FORMS OF  
 CC COLORECTAL CANCER. FAP IS CHARACTERIZED BY ADENOMATOUS POLYPS OF  
 CC THE COLON AND RECTUM, BUT ALSO OF UPPER GASTROINTESTINAL TRACT  
 CC (AMPULLARY, DUODENAL AND GASTRIC ADENOMAS). THIS IS A VICIOUSLY  
 CC PREMALIGNANT DISEASE WITH ONE OR MORE POLYPS PROGRESSING THROUGH  
 CC DYSPLASIA TO MALIGNANCY IN UNTREATED GENE CARRIERS WITH A MEDIAN  
 CC AGE AT DIAGNOSIS OF 40 YEARS.  
 CC -1- DISEASE: APC MUTATIONS HAVE LED TO SOME INTERESTING OBSERVATIONS.  
 CC (1) THE GREAT MAJORITY OF THE MUTATIONS FOUND TO DATE WOULD RESULT  
 CC IN TRUNCATION OF THE APC PRODUCT. (2) ALMOST ALL THE MUTATIONS  
 CC HAVE OCCURRED WITHIN THE FIRST HALF OF THE CODING SEQUENCE, AND  
 CC SOMATIC MUTATIONS IN COLORECTAL TUMORS ARE FURTHER CLUSTERED IN A  
 CC PARTICULAR REGION, CALLED MCR (MUTATION CLUSTER REGION). (3) MOST  
 CC IDENTIFIED POINT MUTATIONS IN THE APC GENE ARE TRANSITIONS FROM  
 CC CYTOSINE TO OTHER NUCLEOTIDES. (4) THE LOCATION OF GERM-LINE  
 CC MUTATIONS TENDS TO CORRELATE WITH THE NUMBER OF COLORECTAL POLYPS  
 CC IN FAP PATIENTS. INACTIVATION OF BOTH ALLELES OF THE APC GENE  
 CC SEEMS TO BE REQUIRED AS AN EARLY EVENT TO DEVELOP MOST ADENOMAS  
 CC AND CARCINOMAS IN THE COLON AND RECTUM AS WELL AS SOME OF THOSE IN  
 CC THE STOMACH.  
 CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.  
 CC -1- DATABASE: NAME-APC; NOTE-Information about APC mutations;  
 CC WWW="http://perso.curie.fr/Thierry.Soussi/APC.html".  
 CC -----

Query Match 2.9% Score 365; DB 1; Length 2843;  
 Best Local Similarity 19.0% Pred No. 1.9e-06;  
 Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;  
 QY 347 PEIRKQREQOE-REFQVQRGAGLSATIRSEH-----EISEIDGLSEQENNEK 395  
 DB 981 PSIESYEDDESKEFCYQYPADLAHKSANHMDNDGELDTPTNYSKYKSYDEQLNSGR 1040  
 QY 396 QMRQLSVIPPMFQAEQRRVFINMGLMEDPMKYKDRQFMVWTDHE--KEIFKDKFI 453  
 DB 1041 Q-----SPSQNERWARPKHIEDEIKQSEQRQSRNOSTTYPVYTESTDDKHL 1087  
 QY 454 QHPKNFGLIASYLERKSVPCVLYYLTKKNYKALVRRNYKRGGRNQOIARPSQEEK 513  
 DB 1088 KQPHFG-----QQECVSPY-----RSRGANGSTNRVSGNHGINQVNSLQOE- 1132  
 QY 514 VEERKEEDKAETK--EKKKDEEKDEKEDSKENTKDKIDGTABETEEREQATRG 571  
 DB 1133 -DDYEDDKPTNYSERYSEEEHHEERTNYSIKYN-EKRVHVDQPIDYSLKYATDIFSS 1190  
 QY 572 R-----KTANSQGRKGRITRSMTEAAAAA-----AAAAATEEPPPPPPP 614  
 DB 1191 QKQSFSSKSSGSSQSKTEHMSSESTSTPSSNAKRONQLHPSSAQSRSQQAATCK 1250  
 QY 615 PEPITSTPEVET-----SRWTEEMEYAKKGLVEHGRN-----WAAIAKM 653  
 DB 1251 VSSINQETIQTQYVEDTPICFSCSSLSLSAEDTEGCTQTQADSANTLQIAETKEK 1310  
 QY 654 VGTKSE-----AOCKNFYNYKRRHNLNLOQHK-----QKTSR 688  
 DB 1311 IGRTSAEDPVSEVPAVSQHPRTKSRLOGSSLSSESARHKAVERFSSGAKSKSGAQTPK 1370  
 QY 689 KPREE-----RQVSCQESVASTVSAQDEDEIASNEENPEPSEVEAVKPSE--DSPEN 740  
 DB 1370 KPREE-----RQVSCQESVASTVSAQDEDEIASNEENPEPSEVEAVKPSE--DSPEN 740

Db 1371 SPPEHYVOETPLMPSRCTSV--SSLDSEFSRSIASSVQSEPCSGMVSGIISPLSDLPDPSGQ 1429  
QY 741 A--TSRGNTEPAVELETETAPSTSLAVPSTKPAEDSVETQVNDSSISAETAQOMDV 798  
Db 1430 TMPSPRSKTPPP--PPQATQKREVPKNAKPTAKRESGPKQAAVNAVO--RVQVLPDA 1485  
QY 799 DQOQSHSAEE--GSCV-----DPPATKADSDVVEVR-----VPEN--HASKVGGDN 840  
Db 1486 DTLHFATETPDGFCSSLSALSLEDPFQK-----DVELRIMPPVQENDNGNETESEQ 1541  
QY 841 TKE--RDLDRASEKVEPRDELVAAQINQAORPEQSDNDSSATCSADEVD----- 890  
Db 1542 PKESNEQKEAEKTI DSKDL-----DSD-----DDDEILEECIIS 1581  
QY 891 ---GEPRQRMFMDKPSLLNPTGSLVSSPLKPNPLD-----LP-----QLHRAAVIP 938  
Db 1582 AMPTKSSRKAKKPAQATASKLPPPVAR-----KPSLPVYKLLPSQNRLOPKQHVSTP 1634  
QY 939 ---PMVSC---TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
Db 1635 GDDMPRYVCEGTPINFSTATSLDIESPPNBLAAGEVGRGAQGEFEKRTIPIEG 1694  
QY 971 AL-----LEEQROEQIDLECRSSSTSPCGTSKSPNREWEVL----- 1007  
Db 1695 RSTDEAOGKTSSTVITPELDDNKAEQDILAEICINSAMPKSKHKPPRVKKIMDOVQOAS 1754  
QY 1008 --QAPHQLTNLPEGVRLPTTRTRPPPLI-----PS 1039  
Db 1755 ASSAPNPK---NQLDGGKKKPTSPVKIPQNTYRTRVRKNADSKNNLNAERFVSDNKDS 1811  
QY 1040 SKTTVAASEKPSF-----IMGSSIS-----QGTGPGYLTSHNOASVYQETPK 1080  
Db 1812 KKQNLKNSKDFNDKLNPNEDRVGRGSAFDSPPHYTIEGTP--YCFSRND----- 1860  
QY 1081 PSVGSISLGLPROQESAKATPLVIKQDEFSRQNSQOPELLVRAQHEGVYRTAGAIQ 1140  
Db 1861 -SLSDFDVDDVLSREKAEELRAKENKESEAKVTSHTE--LTSNQOSA--NKTQAI 1914  
QY 1141 EGSITRGTPTSKISVESIPLSGSIITGTPALPOTGIPTALVK-----GS 1186  
Db 1915 KQINRGOPKPILOKQS-----TFPOSSKDIPRGAATDEKQNFALIENTPVCFSHNS 1968  
QY 1187 ISRM-----PIEDSSPEKREAAASKGHVIEGKSGHI----- 1219  
Db 1969 LSSLDSDIQENNNKENEFIKETEPDQSGPSKQASGAPKSFHVEDTPVCFSRNSSL 2028  
QY 1220 -LSYDN-----IKNAREGTRSPRAHEISLKRSTVESVGNKQGMWRESVPVAPLE 1270  
Db 2029 SLSDSEDDLQECISSAMPKKKPSRLKGDNEKHSRPNMGGILGDLTLDLKDQRPDS 2088  
QY 1271 GLICRALPRGSPHDLKERTVLGSIIMOGTPRATTESFEDGLKYPKQIKRESPIRAFEG 1330  
Db 2089 E---HGLSPDSENFWDKAIQEGANSIVSSLHQAAAAAC-----LSRQASDSDSILSKS 2140  
QY 1331 AITKGPYDGIITTIKEMGRSITHEIPRODILTOESRKTPVEVQVSTPRIIEGSISQGTPIK 1390  
Db 2141 GISLGSFP-----HLTPDQEEKPFTSNKGPRIUK-----PGEKSPLETKKI 2181  
QY 1391 DNNS--GOSATKHNVKSLITGP-----SKLSRGM--PPLIEIVENTIKVBERGYEDVAKGETV 1444  
Db 2182 ESESKIGKGGKVKVYKSLITKVRNSNISEQMKQLOQ---ANMPSISRG----- 2227  
QY 1445 RSRHTSVVSSGPSVLRSLTLEAPKAQLSPGIYDDTSARRTPVSYQNTMSRSGSPMMNRSTD 1504  
Db 2228 -----RTWIIH-----IPGV-----RNSSSSTSPVSKGKGP 2252  
QY 1505 VTIPPKNSTNHERKSTLTPT--QRESIPAK--SPVGVDPVSVSHSPDFPHRGSTAGEVYWS 1562  
Db 2253 LKTPASKSPSEGQATTSPPRGAKPSVKSELSVPARQTSQIGGSKAPSRSGSR----- 2305  
QY 1563 HLPQLDPAFPFHRALDPAANALFQRLQSLFTPGYPSQYQLYAMENTROQLINDYITSQ 1622  
Db 2306 -----DSTPSRPAQOPL-----SRP 2320

QY 1623 MQYNLRLPDVA---RGLSPREQPLGLPYPATRGIIIDLTNMPPTILVHPGCGTSTPPMDRIT 1679  
Db 2321 IQSPGRNISIPGRNGISPPNK-----LSOLPRT--SSPSTASTKSSGSKMS 2365  
QY 1680 YIPCTQITFPFPPVNSMSGPHHTHLAAAASAEERERERE-----KERERE 1727  
Db 2366 Y-----TSPGRMSQONLIT--KOTGLSKNASSIPRESASAKGLNOMNGNGANKVYELS 2417  
QY 1728 RIAAASDLYLRPGSEQGRPGSHGYVRSPSPSVRTQETMLQQRPSVFOQTNGTNGTSVITPL 1787  
Db 2418 RMSSTKSS-----GSESD-----RSEPRVLVRQSTFIKEAPS----- 2449  
QY 1788 DPTAQLRIMPLPAGGPPISQGLPASRYNTAADALALVDAASAPQMDVSKTKRESKHEAA 1847  
Db 2450 -PTLR--RKLEESASFESLS---PSSRPASPTRSQAOPTVLSPSLPDMSL--THSSVQAGG 2503  
QY 1848 --RLEENLRSRSAVSEQOQLEOKTLEVEKRSVQCLYTSSAFFPSGKPKQPHSSVYSEAGK 1905  
Db 2504 WRKLPPNL---SPTIEYNDGRPAKRHDIAI-----SHSESPLRLPNRSCTWKREHSK 2553  
QY 1906 DKGPPPKRSYEEELTRGKTTITAN-----FIDVIITRQIASDKDARERSQSDDS 1957  
Db 2554 HSSSLP--RVSTWRRTGSSSSILSASSESEKAKSEKHNVSISGTSKQKENQVSAKGT 2611  
QY 1958 SSSLSHRYE--TPSDALEVISPASSPAPQOEKLOTYOPEVVKANQAEENDPTROYEGPLHH 2016  
Db 2612 WRKIKENEFSTNSTSTSVSSGATNGAESKTLIYQMAPAVSK---TEDVWVRIEDCPINN 2668  
QY 2017 YRQOESPSQQOQLPP-----SSQAEQMGQVPRHRLITLADHIC 2056  
Db 2669 PR---SGRSTGNTPPVIDSVSEKANPNIKDKNDQAKQNVGNSVPM--RTVGLNRLN 2723  
QY 2057 QITQDFARNQVSSQTP--QOPPTSTFONSALVS--TPVRTKTSNRYSPSPSAQSVHHQR 2114  
Db 2724 SFIOVDAPDQKTEIKPGQNNPVPVSETNESSIVERTPFSSSSSKHSSPSGTVA----- 2778  
QY 2115 PGRSVENLVDRSGRSGKSPERSHVSSEPEPISPPOVPVPH--EKODS 2164  
Db 2779 --AKVTFFNY-----NPSRKSSADSTARSQIP--TPVNNVTKKRD 2818  
  
RESULT 17  
TPR\_HUMAN  
ID TPR\_HUMAN STANDARD; PRT; 2349 AA.  
AC P12270;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NUCLEOPROTEIN TPR.  
GN TPR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93064711; PubMed=1437155;  
RA Mitchell P.J., Cooper C.S.;  
RT "The human tpr gene encodes a protein of 2094 amino acids that has  
RT extensive coiled-coil regions and an acidic C-terminal domain.";  
RL Oncogene 7:2329-2333(1992).  
RN [2]  
RP REVISIONS, AND CHARACTERIZATION  
RX MEDLINE=95096166; PubMed=7798308;  
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,  
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;  
RT "Tpr, a large coiled coil protein whose amino terminus is involved in  
RT activation of oncogenic kinases, is localized to the cytoplasmic  
RT surface of the nuclear pore complex.";  
RL J. Cell Biol. 127:1515-1526(1994).  
RN [3]  
RP SEQUENCE OF 1-142 FROM N.A.

RX MEDLINE-88262257; PubMed-3387099;  
 RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;  
 RT "tpr homologues activate met and raf.";  
 RL Oncogene 2:617-619(1988).  
 CC -1- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE  
 CC COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS  
 CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE  
 CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH  
 CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER  
 CC COMPONENTS, INCLUDING P62.  
 CC -1- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND  
 CC BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.  
 CC -1- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK  
 CC OR RAF GENES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X66397; CA47021.1; -;  
 DR EMBL; Y00672; CA68681.1; -;  
 DR PIR; S00928; S00928.  
 DR MIM; 189940; -;  
 KW Heptad repeat pattern: Coiled coil; Proto-oncogene;  
 KW Chromosomal translocation; Nuclear protein; Transport.  
 FT DOMAIN 78 360  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 422 571  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 575 628  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 758 805  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 834 869  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 934 979  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1004 1064  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1138 1166  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1196 1241  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1262 1304  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1354 1434  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1476 1595  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 527 530  
 FT POLY-SER.  
 FT DOMAIN 1833 1836  
 FT POLY-GLU.  
 FT DOMAIN 1957 1964  
 FT POLY-ASP.  
 FT DOMAIN 2295 2298  
 FT POLY-ASP.  
 SQ SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;

Query Match 2.88; Score 360; DB 1; Length 2349;  
 Best Local Similarity 18.38; Pred. No. 2.3e-06;  
 Matches 484; Conservative 394; Mismatches 966; Indels 806; Gaps 107;

QY 24 SVQYTFNTRHQEQFAVDYRSSHLEVSQASQLQOQ-----QQOQLRRRPSLLSEFHPGSD 80  
 DB 108 AIQSOFTKEELEAEKRDLLRTNERSLQSELEYLTEDVKRLNEKESNTTKGELQLKLD 167  
 QY 81 RPOERTSYEPFHPGSPVDHDSLESKRPRLEQVSDSHFQVSAVLPVHPPLPGLRAS 140  
 DB 168 ELQ-----ASDVSVKREKLEQEKE-----LLHSQNTWLWTE 200  
 QY 141 ADAKKDPAGFKHEAPSPISQPCGDDQNASPKLSKEELIQSDMRVDREIAK-VEQOI 199  
 DB 201 LKTKTDELLALGREGKNEILEKCNLENKKEVSRL--EEQMGLKTSNEHLQKHVEDLL 258  
 QY 200 LKLLK-KQQQ-----LEEAAKPEPEKPPVPPVQKHRSIVQII 239  
 DB 259 TKLKEAKEQQASMEEEKFHNELNAHIKSLNLYKSAADSEAKSNELTRAVEELHKLL----- 314  
 QY 240 YDENRKAEEAAHKIFEGIGPKVELPLYNQPSDTKYVHENIKTNQVMRKL-----ILFFK 294  
 DB 315 -----KEAGEANKAIQDHLLVE-----QSKDQMEKEMLEKIGRLEKELEANDLLSAT 363

QY 295 RRNHARKREQ-----KICORYDQLEAWEK-KVDRIENPRRAK 334  
 DB 364 KRKGAILSEELAAAMSPATAAAVAKIVKPGMKLTLEYLNAVYVETQDQLLEKLENKRINYL 423  
 QY 335 ESKTREYKEKOPPEIRKOREQOORFQVQAGAGLSATIAARSEHEISLIIDLSQENNE 394  
 DB 424 DEIVKE-VEAKAPILKROEYERAKA---VASLSVLEQAMKEIQLR-----QEDTD 473  
 QY 395 KOMQLSV-----IPPMFADABOR----- 414  
 DB 474 KANKOSSVLERNRMEIQVKDLSQLIRVILMELEAARGNHVIRDEEVSSADISSSEVI 533  
 QY 415 ----VKFNMGMLBDMKVIYKD-RQFMNVMTDHEKFKDKFIQHPKNFGLIASYLE-- 467  
 DB 534 SQHLSYRNIEELQOONRLLVALRELGETREREQETSSKITELQLKLESALTELOL 593  
 QY 468 RKS-VPDCVLYLYLTKKENKALVRNYG-----KRGGRNQIAR 507  
 DB 594 RKSQHQHQLVDSIVRQDRMYRILLSSQTGVAIPHAHSLDDVSLASTPKRPSTQSTVST 653  
 QY 508 PSQBEKVEEKEDKAE-----KTEKKKEKKDEEKDEKEDS-----KENTK 549  
 DB 654 PAPVPVISTEALIEAKALKQLOEIFENYKKEAENKIQNEOLEKLOEQVYTDLSQNTK 713  
 QY 550 EKDKIDGTAETE-----EREQATPRGRKTANSQGRKRGKGRITSMNE 592  
 DB 714 ISTOLDFAKRYEMLQDNVEGYRREITSILHERNQ-----KLTTATQ--KQEQIINTMQD 766  
 QY 593 AAAASAAAAATEPPPLPPPPPEISTEPVETSWTEEMEVAKKGLVEHGRNWAATAK 652  
 DB 767 LRGAKEKLAVAEVR-----AENLKKEK-EMKLSEVRLSQOBSLLAORGONLLT 817  
 QY 653 MVGT-----KSEAQCKNFYNY--KRRHLDNLQHQKTSRKPREDVDSOCESVAS 704  
 DB 818 NLQITQILERSEYETKORLSQIEKLEHSHL-----KKLENEVEQRTILTR 867  
 QY 705 TVSAQEDIEDASNEENPEDESEVAVKPSDPSNATSR---GNTEPAVELEPTETAP 761  
 DB 868 NLDVQLDTRQDPTNHLNLTRELLKNAQ--KEIATLKQHLNNEVQVASSQSORTGK 925  
 QY 762 STSPSLAVPSTKPADESIV-----ETOVNDSISAETAQMDVDQOEH--SAEESVCDP 813  
 DB 926 GQ-----PSNKEDVDDLVSQLRQTEQVNDLKERLKTSTSNVEQYQAMVTSLESLAKE 979  
 QY 814 PPATKADSDVDEVVRPN-----HASKVSGDNTKERDLDRASEKVEPREDD 859  
 DB 980 KQVTEEVKNIEVRLKESAEFTQLEKKLMEVEKEKQELQDDKRAIESMEQQLSELAKT 1039  
 QY 860 LVVAQ---QINAQRPESQDNDSSATCSADEVDGEPERQRMFPMDSKPSLLNPTGSIIV 916  
 DB 1040 LSSVQNEVQEQALQRASTALSNEQQAARRDCQEQAKTAVEAQNKYERE----- 1085  
 QY 917 SSPLKPNPLDLPOLQHRAAVTPPVVSVCTPCNIPITGTPVGYALYQRIKAMHESALLEEQ 976  
 DB 1086 ---LMLHAADYEAQAAKEQVSKMAS-----VROH-----LEET 1116  
 QY 977 RORQEOIDLECRSSSPGCTSKSPNREWE---VLQAPAPHOLI---TNLPGEVRLPTTRP 1029  
 DB 1117 TQKESQLLECKAS-----WEERMLKDEVKVCRCEDLEKQNRLLHQ-- 1162  
 QY 1030 TRPPPLIPSSKTTVASSEKPSFIMGSGISQGTPTGTYLTSHNQASYTQETPKPSVGSISLG 1089  
 DB 1163 -----IEKLSKVVASVKEGVQGPLNVSLSEEGKSQE---QILEILRF 1202  
 QY 1090 LPROESAKSATLPYIKOEESPSQNSQOPELLVRAQHEGVVIRGTAGIAQEGSITRGT 1149  
 DB 1203 IRREKETAET-----RFEVAQVESLRYRQVELLER---ELOE----- 1237  
 QY 1150 TSKISVESIPLSGSIQTGTPALQTPGTPEALVKGSI SRMPIEDS-SPEKREEAASK 1208  
 DB 1238 -----LEDSLNAEREKQVQVTAKT 1255  
 QY 1209 HVIIYEGKSGHILSYDNIKNAREGTRSPRTAHEISILKRSYESVEGNIKQCGMRSRPSVAP 1268



FT DOMAIN 1982 1991 POLY-PRO.  
FT DOMAIN 41 1795 4 X 57 AA TYPE A REPEATS.  
FT REPEAT 41 95 1-1.  
FT REPEAT 98 154 1-2.  
FT REPEAT 281 337 1-3.  
FT REPEAT 1740 1795 1-4.  
FT DOMAIN 337 549 2 X TYPE B REPEATS.  
FT REPEAT 337 418 2-1.  
FT REPEAT 476 549 2-2.  
FT DOMAIN 1899 2089 3 X 50 AA TYPE C REPEATS.  
FT REPEAT 1899 1948 3-1.  
FT REPEAT 1965 2014 3-2.  
FT REPEAT 2040 2089 3-3.  
FT CONFLICT 57 57 R -> A (IN REF. 2).  
FT CONFLICT 109 109 Q -> S (IN REF. 2).  
FT CONFLICT 414 414 P -> PPHRGAGNMGPP (IN REF. 2).  
FT CONFLICT 532 532 T -> K (IN REF. 2).  
FT CONFLICT 682 682 Q -> K (IN REF. 2).  
FT CONFLICT 730 730 E -> D (IN REF. 2).  
FT CONFLICT 750 750 L -> R (IN REF. 2).  
FT CONFLICT 834 834 A -> T (IN REF. 2).  
FT CONFLICT 1035 1035 G -> A (IN REF. 2).  
FT CONFLICT 1068 1068 M -> L (IN REF. 2).  
FT CONFLICT 1285 1285 P -> R (IN REF. 2).  
FT CONFLICT 1400 1400 T -> S (IN REF. 2).  
FT CONFLICT 1611 1611 G -> A (IN REF. 2).  
FT CONFLICT 1729 1729 G -> A (IN REF. 2).  
SQ SEQUENCE 2142 AA; 227840 MW; 32DDF16B9B52420A CRC64;

Query Match 2.8%; Score 353; DB 1; Length 2142;  
Best Local Similarity 20.4%; Pred. No. 3.8e-06;  
Matches 439; Conservative 224; Mismatches 833; Indels 656; Gaps 104;

Qy 503 QZIARPS---QEEKVEEKEEDKAEKTEKEEKEEKEDEK-----EKEDSKENTKEKDKI 554  
Db 299 EPVGRPSILKEDNL-----KEFDQDQENDGDWAGAAHEEVDYTEKLFKFSDEE 345

Qy 555 DGTAEETE-----EREQATPRGRKNTANSQGRKGRITRS--MTNEAAAAAASAAAAATE--- 605  
Db 346 DGRSDEGAEGHRDSQSASGEERPPADGKGNSEPTPTKTAWAETSRPTEPEGP 405

Qy 606 -EPPPPPLPP-----PPPISTPEVETSRWTEEEEAQKGLVHGRNWAATAK 652  
Db 406 PAPKPLPPPGDYDPRGGPPCKPAPPEDEAWQRKQSSSEISL--AVERAR----- 456

Qy 653 MVGTKEAQCNFNFTYKRRNLNLLQHKQKTSRPREDRVDSQESVASTVSAQDE 712  
Db 457 -----RRREERMRQEEERRAACAEKL-----KRLDE 483

Qy 713 DIEASNEENPEDSEVAVKPSEDPENATSR---GNTEPAVELEPTTETAPSTSPSLAV 769  
Db 484 KFGADKRLKAEAPAPPA--PSTAPPAPVAPKELPAPPAPPPASAPTEPEPE--EPAQAP 541

Qy 770 P--STKPAEDESVEQVN-----DSISAETAQMDVDQOEHSAE--EGSVCDPPPAT--- 817  
Db 542 PAQSTPTPGVAAAPTLYSGGGSSTSSGSEASVPEPQLPSKEGPEPEVPPPTTPPV 601

Qy 818 -----KADSVDFVRVFNHASKVEGNTKEDLDRASEKVEPRDEDLVAAQINAORPE 872  
Db 602 PKVEFGDGIGFTROPP-----SGLGYPKYOKSL---PPRFQROQQEQQLLQK---QQHQ 651

Qy 873 PQSDNSATSACEDVDGEPEQRMFMDSKPSLLNPTGTSILYSSPL--KENPLDLPLQ 931  
Db 652 WQHQOQSAPTPPPPPQPVITLGVAPAPQAPP---PPPKALYPGALGRPPPPPMNFD 708

Qy 932 HRAAVIPMWSCPTCNPIGTIPVSGYALYQRHIKAMHESALLEBQKQ-----ROEQIDLE 986  
Db 709 PRMMIPPPYD--PRLLQGRPPLEFP-----PGVHPSGLVPRERSDLSLSSEPFDRH 760

Qy 987 CRSSTSPCGTSK--SPNREW-----EVLQAP-----HQLITNLPDGVRLPTTRTPRP 1033  
Db 761 APAMLRERGTPPVDPKLAWGVDFVTATPAEPRLTSLPLRQAADDDKGM--SETPPVPPP 819

Qy 1034 PPLIPSSKTTVASEKPSFIMGSSISQGTPTGTYLASHNQASYQTQETPKPSVGSISLGLPRQ 1093  
Db 820 PPYLAS-----YGFEPENGAPGPI---SRFLEECPRLPWP-----FGS 858

Qy 1094 QESAKSATLPYIKOEFSRSONSQPE--GLLVRAQHEGV-----VRGTAGAIQBS 1143  
Db 859 DEVAKIQTTPPKKEPKKEETAQLTGAAGRLPASRSAGAGPPPPRRRESRTETRMGPRGS 918

Qy 1144 ITRGTPTSKISVESIPLSGSITQGTALPQTGTPTEALVKGSISSRMPIEDSSPKGREE 1203  
Db 919 SRGGIPPEEGAP--PRRAGPIKPPPPPTKVEELPPKPLEQG-----DETAKPPAPD 968

Qy 1204 AASKGHVIEGKSGHILSYDNIKNARECTSPRTAHEISLKRYSVESVGNIKQKMSMES 1263  
Db 969 PLK---ITKGLG-----GPKETPPNGNLSP--APRLRDSYSY-----RVG 1005

Qy 1264 PVSAPLEGLICRALPRGSPHDLKERTVLSGIMOGT-----PRATESFEDGLK 1313  
Db 1006 PTS-----CRGRGRGEYFA-----RGRGFRGTGGRGRGGOANSVATESFEEMM- 1049

Qy 1314 YPKQIKRESPIRAFEGAITKGYDGIITIKEMGRSITHEIPRODILTQESRKTPEVQOS 1373  
Db 1050 -----GV-----EVG-----OGDOTLLLLEA 1066

Qy 1374 TRPIIEGSIQGTPIKFDNNSGSAIKHNKSLITGPSKLSRGMPPLEIVPENIKVVERG 1433  
Db 1067 AMPARHGARGVSMR--KSPSGAGSQAQQAARPMRVIWLLQTRLP----- 1110

Qy 1434 KYEDVKAQETVRSRHTSVVSSGVSRLSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMS 1493  
Db 1111 -----HPRREHSRSPSSRPTSTRPTLHRAPARFTCPGVGESSLP-----EGAIS 1154

Qy 1494 RGSPPMNTSDVT---IPPNKSTNHERKST--LTPTQRESIAPKSPVGVDPVWSHSPDP 1549  
Db 1155 PGRRRREAPPQVCPGWSPPAKSLAPKPTGTLPPSEPLKEKL--IPG-----PLSP 1205

Qy 1550 HHRGTAGEVWWSHLPTOLDPAMPFHRLDPAALAAAYLQROLSPTPGYPVPSQVLYAMENT 1609  
Db 1206 VARGSGNG---GSNVGME--DGERPRRRHGRA-----QQQ-----DKPPFRRLKQE-- 1248

Qy 1610 ROTILNDYITSQOMOVNLRPDVARGLSPREQPLGLPATRGIIDLTMNMPITILVPHPG 1669  
Db 1249 -----RENAARG--SEKPKSLTLPASA-----PG- 1270

Qy 1670 TSTPPMDRITVIGTQITFPFPPPNYSASMP-----GHPTHAAAAAASERERER 1720  
Db 1271 -----PEEAL-----TTVTAPAPRAAKSPDLNQNDSQANEEMETASESDFTSERRG 1321

Qy 1721 EKE-----RERERIAAASDLYLRPGSQPGRPGSHG 1752  
Db 1322 DKEAPPPVLLTPKAVGTPGGGGGAVPGIISAMSRGDLISQRAKDLKSRFSQO--RPGMER 1379

Qy 1753 YVRSPSPSVRTQETMLOQRPSVFOGTNGTSV-----ITPLDPTAQLRIMPLPAGGSIISQ 1808  
Db 1380 QNRRPFGGKA-----GSGSGSGGGGGGGGRTG-----PGRGDKRSM 1419

Qy 1809 LPASRYNTAADALAAALVDAASAFQMDVSKYKESKHEARLEENLRSAVSEQQOLEQ 1868  
Db 1420 SPKNRSRPPEE-----RPFGLPLPPPPSSSAVFLRDQVTHSNPAGI--QOALAQ 1467

Qy 1869 KTLVEKRSVQCLYTSSAFFSGKPKPHSSVVYSEAGDKGPPKRSRYEELRTGKTTIT 1928  
Db 1468 ---LSRQGSVT---APGHPRHKPGP-----POAQPGSPRPTRRYEPQVNSG----- 1511

Qy 1929 AANFDVITITQIASDKDARERGSSDSSSLSHRYETPSDAIEVISPASSPAPPOEK 1988  
Db 1512 -----LSSDPHFEPGPMVRGVG-----TPRDSAGV-----SPFPKRR 1546

Qy 1989 LQ--TYQPEVVKANAQENDPTROYEGPLHYRPOQESPSPOQLPPSSQAEGMGVPTHR 2047  
Db 1547 ERPPRKPELL---QESLPPPHSSGFL---GSKPEGPGQAE-----SRDTG 1587

QY 2048 LITLADHICQIITQDFARN--QVSSOTPOQPTSTFQ-----NSPSALV 2089  
Db 1588 TEALTPHILWNLHTATSRKSYRPTSMEPWMEPLSFEDVAGTQMSQSDSGVDLSGDSQVS 1647  
QY 2090 STPVTKTSNRYSPESQAQSVHH--QRPGRSVSPENLV--DKSRGSRPGKSP-----2137  
Db 1648 SGP-----CSQSSSPDGLKGAAGPPKPRPGGS--SPLNAVPCGPGPSEPPRRPPAPHDG 1702  
QY 2138 ERSVSEPEPIPIPPQVPVHVKODSLLLSQSGAPAEORNDARSPGSIYLPSEFTK 2197  
Db 1703 DRKELPRE--QPLPPG--PCTERSQR-----TDRTGTEPGPIRPSHR--PGP-----PVQFGT 1749  
QY 2198 LENTSPM-----VSKQKQIFRKLNSGGGSDMAAQAQGTIFN--LPATVTTSGSVSS 2249  
Db 1750 SDKSDRLVVGDSILKAKE-----LTAS-----VTEAIPVSRDWELLPSAAASAPQS 1798  
QY 2250 R-----GHISFADPASN--LGLEDIIRKALMGDFDKVEDHGVVM-----2286  
Db 1799 KNLDSGHCVPEPSSGQRLYPEVFGYAGPSSQISGGSHGLSITSQKWLRLRPGTSLHP 1858  
QY 2287 --SOPMGVVPCTANTSVVTSG-----ETREEGDPSPHSGGVCKPKLISKNSRK 2334  
Db 1859 YRSQPLVLPQAPPSALLSGVALKGQFLDFSTQWQATLKLPAAGVLYPPPSFLYSPAF 1918  
QY 2335 SKSPIGOGYLGTERPSVSVSHSEGDYHRO--TPGWAWEDRPSSTGSTQF-----2383  
Db 1919 CPSPLDTSLLQVRQ-----DLPSPSDFYSTPLOGPGSGFLPSGAPAAQOMLLPMDVSQL 1973  
QY 2384 -----PVNPLTMRMLSTPTPTACAPSAVNAQAPHOQNIWEREPAP 2426  
Db 1974 PVNFGSLPPAPPAPPLSLPLVGPALQPPESLAVRPPAPATRVL---PSP 2022

RESULT 19  
YDBJ\_SCHPO  
ID YDBJ\_SCHPO STANDARD; PRT; 661 AA.  
AC Q10369; Q10226;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHETICAL 75.1 KDA PROTEIN C22E12.19 IN CHROMOSOME I.  
GN SPAC22E12.19 OR SPAC2E12.01.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
[1]  
RN  
RP SEQUENCE OF 1-230 FROM N.A.  
RC STRAIN=972;  
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 215-661 FROM N.A.  
RC STRAIN=972;  
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: SOME, TO YEAST YCR33W.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z70043; CAA93906.1; -  
DR EMBL; Z69726; CAA93545.1; -  
DR InterPro: IPR001005; -  
DR Pfam; PF00249; myb\_DNA-binding; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 661 AA; 75096 MW; 59A15D8C990E2C45 CRC64;

Query Match 2.8%; Score 350; DB 1; Length 661;  
Best Local Similarity 21.8%; Pred. No. 1.3e-06;  
Matches 155; Conservative 122; Mismatches 298; Indels 136; Gaps 28;  
QY 224 SPPEV-----EQKHSIVQ-----IYDENRKAEEAHKIFEGIGPKVPELPLY 266  
Db 28 SEPVGMDNEDSDTNTSITQSPNSEKLTDLQESDTRALO-EKYLONIALYALQQLF 86  
QY 267 NQPSDTKYVHENTKTNVMRKLLILFFKRNHARKQREKICQYDQIMAEWKKVDRIE 326  
Db 87 KNVEDSYFYNLNRKFKQDKQITVGMVNRKRRHVLNKKIKRLQSHWKQVLRWEENIARVD 146  
QY 327 N-NPRRKAKESKTEYVEKQPEIRKQREKQERQVQRGAGLSATARSEHSEISIID 385  
Db 147 RLTEIDTKNAKSEPIK-----RSTRKVMSNF-----TAGDIVRSEEFLEILA 192  
QY 386 GLSSEQ---NNEKOMQLSVIPPMFDAEQRRVRFIN-MNGLMEDPMKYKDRQFMVWT 441  
Db 193 KLEQKEKASNVSEASRIATIPPMILSEEVKSOYFNDQSLVTDCKPKFYHFQSMPDIN 252  
QY 442 DHEKEIFKDFIQHPKNGFLIASYLERKSVDPDCVLYIYLLKKNNYKALVRNRYKGRGR 501  
Db 253 EEOHSIFVQOQFILHGKFGKIAEAVPGKSKCEVLHYLTTRTDDYRALV-ASATKTKGR 311  
QY 502 NQIARPSQE-EKVEEKEEDKAEKTEKKEKDEEKEDEKSKENTKEKDKIDGTAE 560  
Db 312 RRKLLPSQGGKKKSGSALWVDIEAADINKTEENINNOFQEA SVTADNMNTWDNT-PS 370  
QY 561 TEEREQATPR-GRKTANSQGRKGRITRSMTNEAAAAAATAEPPLPPPPPPIS 619  
Db 371 VENVESANENVNHNADQMDK---IKSLV-EGNSAYETIEKGAQ-----PDPM 417  
QY 620 ---TEPVETSRWTEEMEVAKKGLVEHGRNWAATAKAVGKSEAKQCNFYFN---YKRRH 673  
Db 418 IDMTDKSETVSGFKHDVVDYTAENEGNNTLLQIKESVHEKTPQDEPMDISODTIKQED 477  
QY 674 NLDNLLQOH--KOKTSRKPREED---VSQESVASTVSAQEDIEASNEENPEDSE 727  
Db 478 YYPEKLEQSHSSKNSISTRKEADAALANLSAVGRSISA-----VDESA 523  
QY 728 VEAVKSPEDSPENA---TSRGNTPEPAVELEP-TTETAPSTSPSLAV-----PSTKPAED 777  
Db 524 HGHLPQWDEKEEALIFSLAQGMNPKMPLTPRRASTGPRPRTFQLTEIDSPNRRASD 583  
QY 778 -----ESVETQVNDISISAETAQMDVDQOQHSAAEESVCDPPATKADSVDEVVRYPEN 831  
Db 584 CITPSISKILKMVSEDASKSRIDELSVDQEH-----T 617  
QY 832 HASKVEGDNT---KERDLDRAS-EKVEPRDEDLVVAQOINNAQRPESQSDND 878  
Db 618 HSSHTTSDINAFPSQSFPFRASHTLAALGEDIV-----ERQSKND 658

RESULT 20  
YKZ6\_CAEEL  
ID YKZ6\_CAEEL STANDARD; PRT; 1018 AA.  
AC P34333;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE HYPOTHETICAL 113.7 KDA PROTEIN C14B9.6 IN CHROMOSOME III.  
GN C14B9.6  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,





CC CRANIOFACIAL DEVELOPMENT THAT OCCURS WITH AN INCIDENCE OF 1/50,000  
CC LIVE BIRTHS. THE CLINICAL FEATURES OF TCS ARE BILATERALLY  
CC SYMMETRICAL AND INCLUDE: (1) ABNORMALITIES OF THE EXTERNAL EARS,  
CC ATRESIA OF THE EXTERNAL EAR CANALS, AND MALFORMATION OF THE MIDDLE  
CC EAR OSSICLES, WHICH MAY RESULT IN CONDUCTIVE HEARING LOSS; (2)  
CC LATERAL DOWNSLOPING OF PALPEBRAL FISSURES, FREQUENTLY WITH  
CC COLOBOMAS OF THE LOWER EYELIDS; (3) HYPOPLASIA OF THE MANDIBLE AND  
CC ZYGOMATIC COMPLEX; (4) CLEFT PALATE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U40847; AAC50903.1; -

DR EMBL; U76366; AAC51181.1; -

DR EMBL; U84664; AAC51185.1; -

DR EMBL; U84640; AAC51185.1; JOINED.

DR EMBL; U84641; AAC51185.1; JOINED.

DR EMBL; U84642; AAC51185.1; JOINED.

DR EMBL; U84643; AAC51185.1; JOINED.

DR EMBL; U84644; AAC51185.1; JOINED.

DR EMBL; U84645; AAC51185.1; JOINED.

DR EMBL; U84646; AAC51185.1; JOINED.

DR EMBL; U84647; AAC51185.1; JOINED.

DR EMBL; U84648; AAC51185.1; JOINED.

DR EMBL; U84649; AAC51185.1; JOINED.

DR EMBL; U84650; AAC51185.1; JOINED.

DR EMBL; U84651; AAC51185.1; JOINED.

DR EMBL; U84652; AAC51185.1; JOINED.

DR EMBL; U84653; AAC51185.1; JOINED.

DR EMBL; U84654; AAC51185.1; JOINED.

DR EMBL; U84655; AAC51185.1; JOINED.

DR EMBL; U84656; AAC51185.1; JOINED.

DR EMBL; U84657; AAC51185.1; JOINED.

DR EMBL; U84658; AAC51185.1; JOINED.

DR EMBL; U84659; AAC51185.1; JOINED.

DR EMBL; U84660; AAC51185.1; JOINED.

DR EMBL; U84661; AAC51185.1; JOINED.

DR EMBL; U84662; AAC51185.1; JOINED.

DR EMBL; U84663; AAC51185.1; JOINED.

DR EMBL; U84664; AAC51185.1; JOINED.

DR EMBL; U84665; AAC51185.1; JOINED.

DR EMBL; U84666; AAC51185.1; JOINED.

DR EMBL; U84667; AAC51185.1; JOINED.

DR EMBL; U84668; AAC51185.1; JOINED.

DR EMBL; U84669; AAC51185.1; JOINED.

DR EMBL; U84670; AAC51185.1; JOINED.

DR EMBL; U84671; AAC51185.1; JOINED.

DR EMBL; U84672; AAC51185.1; JOINED.

DR EMBL; U84673; AAC51185.1; JOINED.

DR EMBL; U84674; AAC51185.1; JOINED.

DR EMBL; U84675; AAC51185.1; JOINED.

DR EMBL; U84676; AAC51185.1; JOINED.

DR EMBL; U84677; AAC51185.1; JOINED.

DR EMBL; U84678; AAC51185.1; JOINED.

DR EMBL; U84679; AAC51185.1; JOINED.

DR EMBL; U84680; AAC51185.1; JOINED.

DR EMBL; U84681; AAC51185.1; JOINED.

DR EMBL; U84682; AAC51185.1; JOINED.

DR EMBL; U84683; AAC51185.1; JOINED.

DR EMBL; U84684; AAC51185.1; JOINED.

DR EMBL; U84685; AAC51185.1; JOINED.

DR EMBL; U84686; AAC51185.1; JOINED.

DR EMBL; U84687; AAC51185.1; JOINED.

DR EMBL; U84688; AAC51185.1; JOINED.

DR EMBL; U84689; AAC51185.1; JOINED.

DR EMBL; U84690; AAC51185.1; JOINED.

CC MIN: 154500; -

CC Disease mutation; Polymorphism.

CC POLY-GLU.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.

CC POLY-SER.









RA Kindler S., Schwanke B., Schulz B., Garner C.C.;  
RT "Complete cDNA sequence encoding rat high and low molecular weight  
MAP2.";  
RL Nucleic Acids Res. 18:2822-2822(1990).  
RN [2]  
RP SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.  
RC STRAIN-WISTAR; TISSUE=Brain;  
RX MEDLINE=91060576; PubMed=2174050;  
RA Kindler S., Schulz B., Goedert M., Garner C.C.;  
RT "Molecular structure of microtubule-associated protein 2b and 2c from  
rat brain.";  
RL J. Biol. Chem. 265:19679-19684(1990).  
RN [3]  
RP SEQUENCE OF 1-151; 1515-1694 AND 1726-1861 FROM N.A.  
RX MEDLINE=90221819; PubMed=2326166;  
RA Doll T., Papandriopoulou A., Matus A.;  
RT "Nucleotide and amino acid sequences of embryonic rat MAP2c.";  
RL Nucleic Acids Res. 18:361-361(1990).  
RN [4]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=89365159; PubMed=2770869;  
RA Papandriopoulou A., Doll T., Tucker R.P., Garner C.C., Matus A.;  
RT "Embryonic MAP2 lacks the cross-linking sidearm sequences and  
dendritic targeting signal of adult MAP2.";  
RL Nature 340:650-652(1989).  
RN [5]  
RP SEQUENCE OF 1695-1725 FROM N.A.  
RX MEDLINE=94110302; PubMed=8282767;  
RA Doll T., Meichsner M., Riederer B.M., Honegger P., Matus A.;  
RT "An isoform of microtubule-associated protein 2 (MAP2) containing  
four repeats of the tubulin-binding motif.";  
RL J. Cell Sci. 106:633-640(1993).  
CC -!- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY  
CC STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO  
CC SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.  
CC -!- ALTERNATIVE PRODUCTS: VARIOUS FORMS OF MAP2 ARE PRODUCED BY  
CC ALTERNATIVE SPLICING OF THE SAME GENE. MAP2C, THE LOW MOLECULAR  
CC FORM OF MAP2, LACKS THE CENTRAL DOMAIN OF MAP2A/B.  
CC -!- DEVELOPMENTAL STAGE: MAP2C IS EXPRESSED DURING EMBRYONIC BRAIN  
CC DEVELOPMENT AND UNTIL POSTNATAL DAY 10. MAP2B IS EXPRESSED  
CC THROUGHOUT BRAIN DEVELOPMENT.  
CC -!- SIMILARITY: CONTAINS 3 OR 4 TAU/MAP REPEATS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X51842; CAA36135.1; -;  
DR EMBL; X17682; CAA35667.1; -;  
DR EMBL; X71487; CAA50588.1; -;  
DR PIR; S07887; S07887.  
DR PIR; S10003; S10003.  
DR PIR; A37981; A37981.  
DR InterPro; IPR001084; -;  
DR Pfam; PF00418; tubulin-binding; 4.  
DR PROSITE; PS00229; TAU\_MAP\_1; 3.  
KW Microtubules; Repeat; Alternative splicing; Calmodulin-binding.  
FT DOMAIN 1454 1474  
FT REPEAT 1664 1694 TAU/MAP MOTIF.  
FT REPEAT 1695 1725 TAU/MAP MOTIF.  
FT REPEAT 1726 1756 TAU/MAP MOTIF.  
FT REPEAT 1757 1788 TAU/MAP MOTIF.  
FT REPEAT 1757 1788 TAU/MAP MOTIF.  
FT VARSPLIC 152 1514 MISSING (IN ISOFORM MAP2C).  
FT VARSPLIC 1695 1725 MISSING (IN ISOFORM WITH 3 TAU/MAP  
FT REPEATS).  
SQ SEQUENCE 1861 AA; 202409 MW; 42DCFF16D21EF54E CRC64;

Query Match

2.6%; Score 332.5; DB 1; Length 1861;

Best Local Similarity 20.0%; Pred. No. 1.9e-05;  
Matches 357; Conservative 237; Mismatches 588; Indels 603; Gaps 94;  
QY 1008 OPAPHQLITNLPEGVRLPTTRTPPPPLIPSSKKTVAASEKSFIMGSGISOGTPTCTYLT 1067  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
DB 119 OPAALPLAAE--ETVNLPP-----PSPPPSPASEQTAALAE-----LL 154  
QY 1068 SHNQASYTQETPKPSVSGSISLGLPRQESAKSATLPYIKQEEFSPRSQNSQOPEGLVRAQ 1127  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 155 TASKMEFFEQQLPS--SFAEPLDKEETEFKMQSKP---GEDFEHAALVPQD----- 202  
QY 1128 HGVVVRGTAGAIQ-----EGSITRTGTPTSKISVESIPSRGSIQTGTPALPOTGIPT 1179  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 203 -----TSKTPQDKKPDQMEGEKSPASFPFAQTGCTNLEDIK-QITEPSITVPSIGLSA 254  
QY 1180 EALV---KGSISRMPIEDSSPEKGREEAASKGHVI-----YECK----- 1215  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 255 EPLAPKQDKQDFIEMPVESKKDEWGLAAPISPGPLTPMREKDVLEIPRWKQKQFDSMP 314  
QY 1216 -----SGHILSYDNIKNAK--EGTRS-----PRTAHEISLKRYSVESVEGNI 1254  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 315 SPFHGGSFTLPDLYKDERVTEGSPFPAPVFQSDDKMSLQDTSGSATSKESKDEEPOK 374  
QY 1255 KQGMSPRESVPASPLEGLICRALPRGSPHSLKERTVLGSGSIMQGTTPRATTFEDGLKY 1314  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 375 DKADKVADVVPVSEATTVL-----GDVHSPAVEGFV--GENISGEKCTTD----- 417  
QY 1315 PKQIKRESPPIRAFEGATKPKPYGITTIKEMGR-SHEIPRODILITQESKRTPE----- 1369  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 418 --QEKKETSTPSVQEPFTLTETEPQ---TKLEETSKVSEETVAKE---BESLKLKDDKAG 469  
QY 1370 VVQSRPIIEGSGISOGTPIKFDNNSGO---SAIKH-----NVKSLITGPSKLSRGMPL 1420  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 470 VIQTS---TEQFSK-----EDQKQEQTIETALQKQSPISLEQAVTDAAMATKLEK 520  
QY 1421 EIVPENI--KVVERGKEDVKAGETVRSRHTSVVSGSVLSRSTLHEAPKAOLSPGI--Y 1476  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 521 TSEPEAVSEKREIOGLFEEDIADK-----SKLEGAGSATVAEV--EMPFYEDKSGMSKY 572  
QY 1477 DTSARRTPVS-----YQNTMSRSGSPMMNRISDVTIPNKSNTNHERKSTLPTQRE 1527  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 573 FETSAKEDVTSTGLGSDYIELSDSRGN--AQESLDTVSPKNQ---QDEKELLAKASQP 627  
QY 1528 SIPAKSPVGVDPVYVSHSPFDPHHRG-STAGEVYVNSHLPTQL--DPAMPFHR--ALDPAA 1582  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 628 SPPA-----HEAGYSTLAGSYSDHPSELPEEPSPSPQERMFIDPKV 669  
QY 1583 AAYLFQRLSPTPGYPQVLYAMENTRTILNDY-----ITSQOMQVNL-----RPDV 1631  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 670 -----YGEKRDLSKKNKDDLTLSRLSLGLGGRSAIEGRSMISNLPMSCLDSI 715  
QY 1632 ARG-----LSPREQPLGLPYPATRGIIDLTN--MPPTILVPHPGTSTPMDRITVI 1681  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 716 ALGFNFGRGHDLSPLASDI---LTNTSGMDGDDYLPPT-----TPAVEKIPCF 762  
QY 1682 P-----GTQIT-----PFRPPY--NSASMSGHPHTLAAA----- 1709  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 763 PIESKEEDKTEQAKVTGGQITQVETSSSEPPAKYKNGTVMAPDLPMLDLAGTRSR 822  
QY 1710 -ASAEERERERERERERERERERERERERERERERERERERERERERERERERERER 1755  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 823 LASVSADAEAARVARKSVSEAVVAESS--TGLPPVADDSQPVKPSDLEDMGVCVENKTYTP 881  
QY 1756 SPSPSVRTQETMLQORPSVFGTNGTSTVITPLDPTAQALRIMPLPAGGSGISOGLPASRYN 1815  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 882 LPSP-VQDSNENLSESGSGFYEGTD--DKVRRDLATDLSLIEVKLAAAG----- 926  
QY 1816 TAADALAALVDA--ASAPQMDVSKTKESKHEA-----ARLENLRSR-----SAAV 1860  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
DB 927 RVKDEFTAEKASPPSSADKSGLSREFQDQRKANDKLDTVLEKSEHVDSEKHAKESSEV 986  
QY 1861 SEQQOLEQKTLVEKRSVQCLYTSSAFPSGKPPQPHSSVVVYSEAGDKGPPPKPSYEELR 1920  
DB : : : : : || : : : : : || : : : : : || : : : : : || : : : : :

Db 987 GDKVFLGLGVTEQTSKALITT-----KETAP-----ERAEGLS 1023

Qy 1921 T-----RCKTTITANFIDVITROIASDKAR--ERGSQS-----DS 1957

Db 1024 SVPEVAEYETTKADQGLDVAAKDDQSPDKVSDFGOMASGMSVDAGKTIKLFVDQ 1083

Qy 1958 SSSLSHRYETPSDAIEVISPASS-----PAPPOEKLOTYQPVVKANQAEINDPTROY 2010

Db 1084 QLTLS---EAPQETDSFMGIESHVHDKAKVSETEVKVAKPDV--HQEAVDKESY 1138

Qy 2011 EGPLHYRQOES-----PSPQOOL-----PPSQAEGMGQVPR 2045

Db 1139 ESSGESHLTKPDGKKEKTSPTSQIODEVALKLSVEIPCPVPVSEAD----- 1190

Qy 2046 HRLTLADHICQIITODFAR--NQVSSOTPOOP--PTSTFQNSALYSTVPRVTKTSNY 2101

Db 1191 ---SSIDEKAE--VOMEFIQLPKEESTETPDIPALPSDVTQPOPEAVVSEPAEVRGEEB- 1244

Qy 2102 SPESQAQSVHHQR-----PGSRVSPENLVDKSRGSRPG----- 2134

Db 1245 --EIEAGEYDKLLFRSDTLQITDLLVPSR--EEFVETCGEKGHVSVVIEDDFI 1299

Qy 2135 ----KSPERSHVSSEPYE--PISPPQV--PVVHEKQDSLLLSQGEPAEQRNDARSP 2185

Db 1300 TVVQTTDEGELGSHSVRFAAPVQPEERRPYPHDEEVLMAEAQAEPKDGSPDAPAT 1359

Qy 2186 GSISYLP-----PSRKLKPAIKYTRTHLSCVRKRTATSGESAQAFKQAKD 1532

Db 1360 PEKEVPSEYKTYDYDKETIDDSIMDADSLWVDTQDDRSILTEQLETIPKEERA 1419

Qy 2204 -----MVKSKKOEIPKLNLSGGG-----DSDMAAAGPOT-----EIFNLPAVTTSG 2245

Db 1420 EKEARRPSLEKHKEKPEK-----TGRGRISTPERREVAKEPSTVSRDEVRKRAVYKA 1475

Qy 2246 SVSS-----RGHSFADPASNLGLEDII-----RKALMGSD-----D 2277

Db 1476 ELAKESEYQAHS---PSRKLKPAIKYTRTHLSCVRKRTATSGESAQAFKQAKD 1532

Qy 2278 KVEDHGVVMS-----QPMGVVPGTANTSVTSGETREEDGDPSPHSGVCKPKLISK 2329

Db 1533 KWTG-DITKSEPKRSLLRPSILPRRG-----VSGD--RENSFSLNS--ISS 1578

Qy 2330 SNRSKSPFPGQYLGTRPSSVSSV-----HSEGDYHROTPGNAWEDRPSSTGSTOPPY 2385

Db 1579 ARRTTRSEPIRRAGSGTSTPTPGSTAITPGTPSYSSRTPG-----TPGTPSYPR 1630

Qy 2386 NPLNMR-----MLSTPTPIACAPS--AVNQAAPHQON 2417

Db 1631 TPGTPKGLVPEKKAIVAIIRTPKSPATPKQLRLINQPLDLKN 1675

RESULT 25

AKAC\_HUMAN

ID AKAC\_HUMAN STANDARD; PRT; 1781 AA.

AC Q02952; Q99970; Q00498; Q00310;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE A-KINASE ANCHOR PROTEIN 12 (A-KINASE ANCHOR PROTEIN 250 KDA) (AKAP 250) (MYASTHENIA GRAVIS AUTOANTIGEN GRAVIN).

DE AKAP12 OR AKAP250.

GN Homo sapiens (Human).

OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.

OX NCBI\_TaxID=9606;

[1]

RN SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Heart;

RX MEDLINE=97153077; PubMed=9000000;

RA Naurel J.B., Klauk T.M., Langeberg L.K., Scott J.D.;

RT "Gravin, an autoantigen recognized by serum from myasthenia gravis patients, is a kinase scaffold protein.";

RL Curr. Biol. 7:52-62(1997).

[2]

SEQUENCE FROM N.A. (ISOFORM 2).

TISSUE=Umbilical vein endothelial cells;

MEDLINE=98269042; PubMed=9604001;

RA Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;

RT "Changes of gene expression by lysophosphatidylcholine in vascular endothelial cells: 12 up-regulated distinct genes including 5 cell growth-related, 3 thrombosis-related, and 4 others.";

RL J. Biochem. 123:1119-1126(1998).

[3]

SEQUENCE OF 43-1781 FROM N.A.

TISSUE=Umbilical vein endothelial cells;

Bowditch R.D., Ginsberg M.H.;

"Sequence of gravin cDNA isolated from a human umbilical vein endothelial cell library.";

Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

[4]

SEQUENCE OF 1477-1781 FROM N.A.

TISSUE=Umbilical vein endothelial cells;

MEDLINE=92395179; PubMed=152245;

Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R., Lindstrom J., Ginsberg M.H.;

"Molecular cloning and preliminary characterization of a novel cytoplasmic antigen recognized by myasthenia gravis sera.";

J. Clin. Invest. 90:992-999(1992).

-1- FUNCTION: ANCHORING PROTEIN THAT MEDIATES THE SUBCELLULAR COMPARTMENTATION OF PROTEIN KINASE (PKA) AND PROTEIN KINASE C (PKC).

-1- SUBUNIT: BINDS TO DIMERIC RII-ALPHA REGULATORY SUBUNIT OF PKA.

-1- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY BE PART OF THE CORTICAL CYTOSKELETON.

-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; MAY BE PRODUCED BY ALTERNATIVE SPLICING.

-1- TISSUE SPECIFICITY: EXPRESSED IN ENDOTHELIAL CELLS, CULTURED FIBROBLASTS AND OSTEOSARCOMA, BUT NOT IN PLATELETS, LEUKOCYTES, MONOCYTIC CELL LINES OR PERIPHERAL BLOOD CELLS.

-1- INDUCTION: ACTIVATED BY LYSPHOSPHATIDYLCHOLINE (LYSOPC).

-1- DOMAIN: POLYBASIC REGIONS LOCATED BETWEEN RESIDUES 265 AND 556 ARE INVOLVED IN BINDING PKC.

-1- DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY PATIENTS WITH MYASTHENIA GRAVIS (MG).

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; U81607; AAC51366.1; -

EMBL; AF001504; AAB58938.1; -

EMBL; AB003476; BAAL1927.1; -

EMBL; M96322; AAA35931.1; -

MIN; 604698; -

PIR; A43922; A43922.

InterPro: IPR001573; -

InterPro: IPR001899; -

Antigen; Alternative splicing.

DOMAIN 603 633 AKAP 1.

DOMAIN 752 782 AKAP 2.

DOMAIN 797 827 AKAP 3.

DOMAIN 98 101 POLY-GLU.

DOMAIN 265 556 INVOLVED IN PKC-BINDING (PROBABLE).

DOMAIN 1540 1553 MISSING (IN ISOFORM 2).

VARSPIC 1 98 RII-BINDING (PROBABLE).

VARSPIC 99 106 EEEVITE -> MLGITIT (IN ISOFORM 2).

CONFLICT 117 117 E -> K (IN REF. 2).

CONFLICT 142 144 NRN -> TPEI (IN REF. 2 AND 3).

CONFLICT 215 215 Q -> K (IN REF. 2).

CONFLICT 448 448 G -> E (IN REF. 2 AND 3).

CONFLICT 694 694 R -> G (IN REF. 2 AND 3).

CONFLICT 867 867 G -> S (IN REF. 2 AND 3).











OX NCBI\_TaxID-10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE-Glial tumor;  
 RX MEDLINE-91268156; PubMed-2050743;  
 RA Wiche G., Becker B., Luber K., Weitzer G., Castanon M.J.,  
 RT Hauptmann R., Stratowa C., Stewart M.;  
 FT "Cloning and sequencing of rat plectin indicates a 466-KD polypeptide  
 chain with a three-domain structure based on a central alpha-helical  
 coiled coil.";  
 RL J. Cell Biol. 114:83-99(1991).  
 RN [2]  
 RP REVISIONS.  
 RC TISSUE-Glial tumor;  
 RX MEDLINE-96210632; PubMed-8633055;  
 RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;  
 RT "Human plectin: organization of the gene, sequence analysis, and  
 chromosome localization (9q24).";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).  
 RN [3]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4), AND TISSUE SPECIFICITY.  
 RC TISSUE-Glial tumor;  
 RX MEDLINE-97321050; PubMed-9177781;  
 RA Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,  
 RT Wiche G.;  
 FT "Plectin transcript diversity: identification and tissue distribution  
 of variants with distinct first coding exons and rodless isoforms.";  
 RL Genomics 42:115-125(1997).  
 CC -1- FUNCTION: PLECTIN IS PROPOSED TO PLAY A ROLE IN CROSS-LINKING  
 INTERMEDIATE FILAMENTS, TO INTER-LINK INTERMEDIATE FILAMENTS  
 WITH MICROTUBULES AND MICROFILAMENTS, AND TO ANCHOR INTERMEDIATE  
 FILAMENTS TO MEMBRANES, TO THE PLASMA MEMBRANE AS WELL AS TO THE  
 NUCLEAR MEMBRANE. PLECTINS HAVE THE ABILITY TO SELF-ASSOCIATE AND  
 FORM NETWORKS THAT STABILIZE THE CYTOPLASM.  
 CC -1- SUBUNIT: A TETRAMERIC STRUCTURE IS PROPOSED WHERE THE TWO CHAINS  
 IN ONE MOLECULE ARE PARALLEL TO ONE ANOTHER AND THE TWO MOLECULES  
 ARE ARRANGED ANTIPARALLEL AND OVERLAP TO A GREAT EXTENT.  
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE  
 PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN  
 SKELETAL MUSCLE AND LOWEST IN THYMUS.  
 CC -1- SIMILARITY: TO DESMOPLAKIN AND TO BULLOUS PEMPHIGOID ANTIGEN.  
 CC -1- CAUTION: ISOFORM 4 IS A FRAGMENT AT THE N-TERMINUS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59601; CAA42169.1; -;  
 DR EMBL; U96274; AAC53209.1; -;  
 DR EMBL; U96275; AAC53210.1; -;  
 DR EMBL; U96276; AAC53211.1; -;  
 DR PIR; A39638; A39638.  
 DR PIR; S21876; S21876.  
 DR HSSP; Q01082; 1AA2.  
 DR InterPro; IPR001101; -;  
 DR InterPro; IPR001715; -;  
 DR Pfam; PF00307; CH; 2.  
 DR Pfam; PF00681; Plectin\_repeat; 21.  
 DR PROSITE; PS00019; ACTININ.1; FALSE\_NEG.  
 DR PROSITE; PS00020; ACTININ.2; FALSE\_NEG.  
 DR PROSITE; PS00021; CH; 2.  
 DR Coiled coil; Repeat; Structural protein; Cytoskeleton;  
 KW Actin-binding; Alternative splicing.  
 FT DOMAIN 1 400 ACTIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 1 71246 GLOBULAR.  
 FT DOMAIN 71247 2648 CENTRAL FIBROUS ROD DOMAIN R.  
 FT DOMAIN 71247 1474 R1 COILED COIL.  
 FT DOMAIN 1475 2648 R2 COILED COIL.

|    |          |          |            |                                                     |
|----|----------|----------|------------|-----------------------------------------------------|
| FT | DOMAIN   | 1475     | 1686       | R2-A SUBDOMAIN (POTENTIAL).                         |
| FT | DOMAIN   | 1727     | 1927       | R2-B SUBDOMAIN (POTENTIAL).                         |
| FT | DOMAIN   | 1967     | 2124       | R2-C SUBDOMAIN (POTENTIAL).                         |
| FT | DOMAIN   | 2298     | 2489       | R2-D SUBDOMAIN (POTENTIAL).                         |
| FT | DOMAIN   | 2490     | 2648       | R2-E SUBDOMAIN (POTENTIAL).                         |
| FT | DOMAIN   | 2786     | 4620       | GLOBULAR.                                           |
| FT | DOMAIN   | 787      | 930        | COILED COIL (POTENTIAL).                            |
| FT | DOMAIN   | 1105     | 1140       | COILED COIL (POTENTIAL).                            |
| FT | DOMAIN   | 1241     | 1307       | COILED COIL (POTENTIAL).                            |
| FT | DOMAIN   | 1378     | 1434       | COILED COIL (POTENTIAL).                            |
| FT | DOMAIN   | 1472     | 2760       | COILED COIL (POTENTIAL).                            |
| FT | DOMAIN   | 1727     | 1897       | INTERACTION WITH VIMENTIN AND LAMIN B.              |
| FT | DOMAIN   | 1727     | 1897       | 6 X TANDEM REPEATS OF MOTIF A.                      |
| FT | REPEAT   | 2786     | 3113       | MOTIF A (APPROXIMATE).                              |
| FT | REPEAT   | 3114     | 3444       | MOTIF A (APPROXIMATE).                              |
| FT | REPEAT   | 3445     | 3779       | MOTIF A (APPROXIMATE).                              |
| FT | REPEAT   | 3780     | 4024       | MOTIF A (APPROXIMATE).                              |
| FT | REPEAT   | 4025     | 4367       | MOTIF A (APPROXIMATE).                              |
| FT | REPEAT   | 4368     | 4620       | MOTIF A (APPROXIMATE).                              |
| FT | DOMAIN   | 2845     | 4597       | 6 X 19 AA REPEATS.                                  |
| FT | REPEAT   | 2845     | 3015       | 1.                                                  |
| FT | REPEAT   | 3173     | 3343       | 2.                                                  |
| FT | REPEAT   | 3504     | 3674       | 3.                                                  |
| FT | REPEAT   | 3839     | 4008       | 4.                                                  |
| FT | REPEAT   | 4082     | 4252       | 5.                                                  |
| FT | REPEAT   | 4427     | 4597       | 6.                                                  |
| FT | DOMAIN   | 4628     | 4643       | 4 X 4 AA TANDEM REPEATS OF G-S-R-X.                 |
| FT | DOMAIN   | 4645     | 4687       | SER-RICH.                                           |
| FT | VARSPLIC | 1        | 37         | MSQRLRPVPEGLSGKRTSSDNLVLAIVLRASEGKK (IN ISOFORM 2). |
| FT | VARSPLIC | 1        | 37         | MSQRLRPVPEGLSGKRTSSDNLVLAIVLRASEGKK (IN ISOFORM 2). |
| FT | VARSPLIC | 1        | 51         | MSQRLRPVPEGLSGKRTSSDNLVLAIVLRASEGKK (IN ISOFORM 3). |
| FT | VARSPLIC | 38       | 180        | MSQRLRPVPEGLSGKRTSSDNLVLAIVLRASEGKK (IN ISOFORM 4). |
| FT | VARSPLIC | 52       | 180        | MISSING (IN ISOFORM 2 AND ISOFORM 3).               |
| FT | VARSPLIC | 52       | 180        | MISSING (IN ISOFORM 4).                             |
| SQ | SEQUENCE | 4687 AA; | 533527 MW; | 9966CA71B929751 CRC64;                              |

Query Match 2.5%; Score 317; DB 1; Length 4687;  
 Best Local Similarity 18.3%; Pred. No. 0.0002;  
 Matches 545; Conservative 398; Mismatches 1096; Indels 932; Gaps 129;

|    |      |                                                                  |
|----|------|------------------------------------------------------------------|
| Qy | 10   | QCAFSTESRPPPHSVQVTFNTRHQEF-----AVPDRSSHLEVSQASQLLQOQQQ 62        |
| Db | 1546 | EAADVDAQOQK---RSIQEELQHLRQSSEAEIQAKQAQVEAAERSMRIEEIRVVRQLQE 1602 |
| Qy | 63   | QQLRRRPSLLSEFHGSDRPOE---RRTSVPPFHPGPSVDHDSLESKRPRLEQVSDSH 118    |
| Db | 1603 | TERQRGGADELQALRAAEAEAKKQAOE-----EAERLRQVQDESQ 1648               |
| Qy | 119  | FORVSAAYPLVPLHPLPEGLRASADAKDPFAGKGHEAPSSPTSGQPCGDQNASPSKLSK 178  |
| Db | 1649 | RKRQAEAEAL------RVKAEAEARE-----KQAL-----QALDELKQA 1685           |
| Qy | 179  | BELTQSDMRVDRETAKVEQIILKKKKOQLEEEAEAKPEPEKPVSPPPVEQKRRSIVQ 237    |
| Db | 1686 | EEAEERLQCAEAEARARQVVALETAQRAEAEVQLQKRPFAEKTAQLERTLQEEHVTVQ 1745  |
| Qy | 238  | IYDENRKKAEAEHAKIFEGLPKVELPLYNPSDFKVVHENIKTNQVNRKKLILFFKRN 297    |
| Db | 1746 | LREAEERRAQQA-----EAERAEAEELERWQLKANEALRLR----- 1786              |
| Qy | 298  | HARKOREQKICQRYDQLMEAEKKVDRIENPNRRKAKES-----KTREYVEKOFPEIRKQ 353  |
| Db | 1787 | QAEVAQOQKSLAQADA-----EKQKEAEERARRRGKAEQAVRORELAEQ-----ELEKQ 1838 |
| Qy | 354  | E-----QOERFORVQRGAG-----LSATITARSEHS----- 381                    |
| Db | 1839 | QLTEGTAQORLAAEQELIRLRAETTEQGHQRLLEELARLQHEATAATQKQLEAEALA 1898   |

|    |      |                                                                |      |
|----|------|----------------------------------------------------------------|------|
| Qy | 382  | -----EIIDGLSEQNENKQMRLSVIPPMFMDAEQRVXFI-----NMNGLMEDPMKY       | 430  |
| Dd | 1899 | KVRAEMEVL--LASKARAEESRSTSEKSQBLEAGRFRELAEEAARLALAAEARR-        | 1955 |
| Qy | 431  | YKORQFNWVTDEKEFTFKDKFIQHPKNFGLIASYL-----ERKSPVDCLVLYLLTKN      | 484  |
| Dd | 1956 | -----HRELAEDAARQADGVITEKLAIASEATRLKTAEIALKEKAEN                | 2003 |
| Qy | 485  | ENYKALVRNRNGRRGRNOQIA-----RPSQBEKVEEK-----BDKAETKEKK           | 528  |
| Dd | 2004 | EURLRLADEAFORRRLEBQAQHKADIERLAQLRKASELERQKGLVEDTLRQRQV         | 2063 |
| Qy | 529  | EBEKKDDEEKDEK-EDSKENTKERDKIDGTAEET---EREQATPRGRTANSQGRKK       | 582  |
| Dd | 2064 | EEEIMALKASFKAAGAKGAELELGRTNSNAEDTMRSELAEFAARQOLAAEEBQRR        | 2123 |
| Qy | 583  | ----GRITRSMTNEAAAASAAAAATEPPPLPPPPPPISTEPEVTSRWTEE-EMEVAK      | 637  |
| Dd | 2124 | REAERVORSIAEEEEAQRKVALEE-----VERLKAYVEARRLRERAQEQSAR           | 2174 |
| Qy | 638  | KLIVEHGRNWAAITAKMVGYSQAQCKNFYNTKRRHNLDNLLOHK-----Q             | 684  |
| Dd | 2175 | QLQLAQ---EAAQKRLQAEKAHA---FVVQOREELQQOTLQOEONMLERLSRSEAAR      | 2227 |
| Qy | 685  | KTSRKPRERDVSOCESVASTVSAODEDIEASNEENPEDSEYEA-V-KPSEDSPENAT      | 742  |
| Dd | 2228 | RAAEAEAREQEAREAAOSRQVFEARELKQSAEEQAOQAQAAAEKLRKEAQEAA          | 2287 |
| Qy | 743  | SRGNETPA-----VELEPTTETAPST-----SPLAVPSTKPADESIVETOVD           | 786  |
| Dd | 2288 | RRAQAEQAALLKQQAADAEMEHKKFAEOTLRKAQVQBELTTLRLQLTEETHOKSLIDE     | 2347 |
| Qy | 787  | SISAETAEQMDVDOQESHSAEGSVCDPPATKADSDVDEVVRPNVSHASKVEGDN----     | 840  |
| Dd | 2348 | ELQRLKAEVTEAARQRSQVE-----ELFSVRVOMEELGKLRKAEIENRALILR          | 2397 |
| Qy | 841  | ---TKERDLDRASEKVEPRDEDL---VVAQINAQRPESQSD-----876              |      |
| Dd | 2398 | DKONTORFLEEEAKMKQVAEEAARLSVAQAEARLQLABEDLAQORALAEKLMKEKMQ      | 2457 |
| Qy | 877  | -----NDSSATCSAD-----EDVDG-----EPERQRMFP                        | 900  |
| Dd | 2458 | AVQEAIRLKAEBLLQQQKELAQEARRIQADKEQMAQLVEETQGFGORTLEABERQOLE     | 2517 |
| Qy | 901  | MDSKPSSLNPTGSIIVSPKPNPLDLPOQHRAAVIPPVWSCTPCNPICPTPVSGVALY      | 960  |
| Dd | 2518 | MSAE-----AERLKRMAEMSRAQARAEEDAQFRKQAEI-----GEKLH               | 2558 |
| Qy | 961  | QRHIKAMHESAL---LEBORQROEQIDLCECRSSTSFCQTSKSP-NREWELVOLAPHOLIT  | 1016 |
| Dd | 2559 | RTELATOQKTVLTQTLEIQROSDDQDAERLRAIAELEREKEKLUKOEAKLQLKSEMQT     | 2618 |
| Qy | 1017 | NLPEGVRLPTRTPTRPPPLIPSSKTTVASEPSPFMGGSI-----SOCTPG             | 1063 |
| Dd | 2619 | VQEQIILQET-----QALQKSLFSEKDSLQORERIEQEKALEQLFODEFVAKA          | 2667 |
| Qy | 1064 | TYLTSHNQASYTO--ETPKPVSIGSISLGPLROESAQSATLPYIKOBEFPSRSONSQPEG   | 1121 |
| Dd | 2668 | KQLQEEQORQQQMEQEKQELVASMEARRRQREAEQVRR---KQBELQRLEQOOROQOE     | 2724 |
| Qy | 1122 | LIVRAQHEGVVRCAGIAIOBSITRGTPTSKISVESISFLRGSITOGTTPALPQTGIPTFA   | 1181 |
| Dd | 2725 | KLAAENQRL-----RELQRLDEEHRALAHSSEIATSQAATKALPN---GRDA           | 2772 |
| Qy | 1182 | LVBGSI SRMP-----IEDSSPEKGREEAASKGHVIYEGKSGHILSYDNINKNAREGRSRP  | 1235 |
| Dd | 2773 | LDGPSMEAREPYTFEGLRQVKVAQQLQERAG-----ILSMEEQLRUTQG---H          | 2816 |
| Qy | 1236 | RTAHEISLKRYSVESBGNITKQCMNSWRSPVSNAPLEGL-ICRALPRG--SPHSD--LKERT | 1290 |
| Dd | 2817 | TTVAEITQR----EDVRHYLKGSSLAGLLKPTNEKLSVYTALQRLSLSGCTAILLEAQ     | 2873 |

|    |      |                                                                |      |
|----|------|----------------------------------------------------------------|------|
| QY | 1291 | VLSGSIM---QGTPRATTSFEDGLKYKPKQIKRESPTIRAEIGAITGCK--PYDGIITTIKE | 134  |
| Db | 2874 | AASGFLLDPVNRRLTVNEAKVGVPGLHHK-----LLSAERAVGTGYKDPYTGEI---      | 2926 |
| QY | 1347 | MGRSIIHELPRODILTOE-SRKTEPEVVQSRTPLIEGSIQSOTPIKFDDNNSGOSAIKNHVS | 1405 |
| Db | 2927 | --SLFOAMKKDLIVRDHGIRLLEAQIATGGIIDPVHSHRPVDVAVQRGY--FDEEMNR     | 2981 |
| QY | 1406 | LITGPSKLRSKM--PPLEIVPENIKYVERGKYEDKAGETVRSRHTSVSSGSPVLSTL      | 1463 |
| Db | 2982 | VLADPSDDTKGFFDPNTH--ENLTVIQ-----LLERCVEDPETGLRLPLTD            | 3026 |
| QY | 1464 | HEAPKAOLSGIYDDTSAR-----RTPVS-----YONTHSRCSPM                   | 1498 |
| Db | 3027 | KAAKGSEL---VYTDPEAROVFEKATVSAPFKCGFGKTWTWIEINSEYFTAEORDLLR     | 3083 |
| QY | 1499 | MNRTSDVTIPP-----NKSTHERKSTLT-PTQRESIPAKSPGVDPVYVSHSFDPDH       | 1551 |
| Db | 3084 | QFRTRITVEKIIVITVEEHERRKGQLCFSELALYPAAELLD--SGVISHEVYOQLQ       | 3141 |
| QY | 1552 | RG-----STAG-----EYVWSHLPTQ                                     | 1567 |
| Db | 3142 | RGRSVREVAEADEVQALRGTSVIAGVWLEAGOKLSIYEALRRDLLQPEVAVALLEAQ      | 3201 |
| QY | 1568 | ----LDAMPFHRLDPAAMAAYLFORQLSPTPGPYSOYLAYMENTROTILANDYITSQ      | 1621 |
| Db | 3202 | ACTGHIDPAT SARLVDEAVRAGLV-----PEMHEKILLSAENVGYDRPYSGQ          | 3252 |
| QY | 1622 | QMVNLRPDVARGLSPREOPLGL--PYPATRGIDLTNNPPTLVPHPGGHSPTPMDRIT      | 1679 |
| Db | 3253 | --SVSLFOALKKGLIPREOGLRLDAQLSTGGIVD-----PSKSHRPVLD-V A          | 3297 |
| QY | 1680 | YIPQTQITEPPRPYNASASPGHPH-----LAAA                              | 1709 |
| Db | 3298 | YARG-----YLDKETNALYSPRDDARVYLDPTREPVIYSOLOQRCSDQLTGLSLPLS      | 3353 |
| QY | 1710 | ASAEREREREKE-RERERIAAA-----SSDLYLRPGSEQ----                    | 1744 |
| Db | 3354 | EKAVRAROEVIYSELOARETLEKAKVEVPVGGFKGRALTVMELISSEYFTEBQREQELLRQ  | 3413 |
| QY | 1745 | -PCRPBGSHGVSPSPSYVRTOMLQORPSVFQCTNGT SVITPLDPTAQLRI MPLPAGG    | 1802 |
| Db | 3414 | FRTQKVIVEIKILLITYEEVEETQBERLS-FSGLRA-----PVFASE                | 3456 |
| QY | 1803 | PSISQGLPASRYNTAADALAALYD-----AAASAPQMDVSKTESKHEAAR             | 1848 |
| Db | 3457 | LLASKILSRTOFEOQLDGKTSVKDLSEVGSVRTLLQSGCLAGIYLEDSKEKVTIYEAMR    | 3516 |
| QY | 1849 | LEENLRSRSAVSEOOOLEKITLEKRSVOCLYTSAPPSG--KPQHSSHVVSYSEAGKD      | 1906 |
| Db | 3517 | -RGLLRASTATLLEQAATGFL-VDPVRNQRLYVHEAVKAGVGPVELHEKILLSAEKAVT    | 3574 |
| QY | 1907 | KGPPPKS-----RYEBELRTRG-----KTTITAANFDIVITROIASDKDARBEG-        | 1951 |
| Db | 3575 | GKDPYSGSIVLSFOAMKKGVLRLDHAIRLLEAQIATGGIIDPVHSHRLPVDV-A YORG Y  | 3633 |
| QY | 1952 | -----SQSDSSSSL---SSHRYETPSDAIE--VISPASS-----PAPPQBKLQTYQ       | 1993 |
| Db | 3634 | FDEEMNVLPADPSDDTKGFFDPNTHENLYTLQLLERCVEDPETGLRLPLRGAEK-----    | 3688 |
| QY | 1994 | PEVVKANO--AENDPTRQYE-----GPLHH-----YRPQESPSPOOO---LPSS         | 2034 |
| Db | 3689 | TEVVETTQVYTEETRAEFETQIDIPGGSGHSGSMLEWFMQSDMIPEQARLAMDF         | 3748 |
| QY | 2035 | QAEGMGVQVPTHRLITLADHI--CQITO-----DFARNQVSSQ-----               | 2071 |
| Db | 3749 | QA---GRVTKERMIIIITIEIKETEIROONLASVDYVRRRLTAEDLYEARII SLEYTNL   | 3805 |
| QY | 2072 | -----TPQPPTST-FQNSPALYSTVPVRTYSN                               | 2099 |
| Db | 3806 | FREGTKSLREVLEMESAWRYLXTGVSAGVLYLPGSRQTLTIYQALKKGLLSAEVA-----   | 3860 |
| QY | 2100 | RYSPESAQSVHHQRCP--GSRVSPENLVDKSRGRPG-----KSPERSHVS-SEP         | 2146 |



```
FT CARBOHYD 951 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2054 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2362 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3342 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 349 MISSING (IN ISOFORM V1).
FT VARSPLIC 1309 MISSING (IN ISOFORM V2).
FT VARSPLIC 349 MISSING (IN ISOFORM V3).
FT CONFLICT 348 P -> R (IN REF. 2).
SQ SEQUENCE 3358 AA; 366938 MW; 071B80026BC0762D CRC64;

Query Match 2.5%; Score 314.5; DB 1; Length 3358;
Best Local Similarity 19.6%; Pred. No. 0.00017;
Matches 457; Conservative 282; Mismatches 803; Indels 795; Gaps 119;

Qy 582 KGRITRSMNEA-RAASAAAAAATEEP--PPPLPPPEPISPE-PVETSRW-----T 629
Db 349 KONISEATTIEMNLAETSSPLSKPEHVPDRATVPVPLATELPIFTTHPPAGNIYNS 408
Qy 630 EEMEVAKGLVEGRNWAALAKMVGTKSEAQCKNFYFKRRHNLNLLQOHKQKTSRK 689
Db 409 EOKSVVYSQ-----AITGRLAYESPITTRNTINSWDLNDSL-----ASGSG 449
Qy 690 PREERDVSC---RSVASTVSAQDEDEDIEASNEENPDSVEAVKPSDESDP----- 738
Db 450 PLGMPDISEIKEELRSTVTISQHATGSOAVITDQTHESVSQIEQIEVGPLVTSMEIT 509
Qy 739 -----ENATSRGTEPAVELEPTE--TAPSTSPSLAVPS----- 771
Db 510 NHISLKEPKNKTPYESTE--VLEHTTEMPTVSASPELATTSYFTLREDDREDRTL 567
Qy 772 -----TKPADESVETQVND--SISAETAEQMD-----VDQOEHS 804
Db 568 TVRSQDSTRVFSQIPVITVSKTSED--TYSQGLDLESISTITMLGTRSLDKKEP 626
Qy 805 AEGSVCD-----PPATKADS----VDVEVRVPENHASKVGDNT-----KEROLDRASE 851
Db 627 KTKGKVTEDERGQSQPTTFPSQHLTEVELLPYSQDTSVEGISTVIYPSLQTDVTOGRE 686
Qy 852 KVE-PRDE---DLVVAQOINARPE-----PQSONDSATCSADEV 889
Db 687 RTETPRPELKDPYTVDEI-----PEKVTKDFIGKTEVFSGMPLSTSSSESVERTESV 742
Qy 890 D-----CEPERQR-----MFPMDSKPSLLNPTGSTL----- 915
Db 743 SPALTIELTKPTEARDVEEMTILTRLETDVTKSDKDVTRVHLTHSTLANVEVTVSKWP 802
Qy 916 ---VSSPLKPNPLDLPOLQHRA-AVIPP-----VSCTPCNIPIGTPVSG-YALIQRHI 964
Db 803 GDEDNSTSKP---LPSTEHAFTKLPVPLSTINGCKDKKEIPSFDTGGGEYTLFPDGT 858
Qy 965 KAMESALLEQRORQOQIDLECRSSTSPCGTSKSPRE-----WEVLQAPAPQLITNL 1018
Db 859 PKPLEK---VSEEDLASGELATVTFHTSTSIGSAEKSASGEPTTGDRFLPTSTEDOVINAT 916
Qy 1019 PEGVRL-----PTTRTPRPPPLIPSKTTVASEKPSFTMGSSISQGTPTGYL-TSHNOAS 1073
Db 917 AEGSALGEDTEASKPLTGPFFVHTSDV-----EELAFVNYSSTOE--PTIYVDISH---- 966
Qy 1074 YQETPKPSVSGISLGLPROQESAKSATPLVIKQDEFSPPRSQSOPEGLVRAHQEGVVR 1133
Db 967 -----TSPLSI-IPKTEWSVLETSPV-LEDEILGKSDQDILEQTHLEATMSPGALR 1015
Qy 1134 GTA---GAIOEGSITRGTPTSKISVESI-----PSLRGSITOGT---P 1170
Db 1016 TTGVSQGETQEEPOTGSPFFTFSTAVMAKETTAEEGEGSTYTPS--EGLMTGSRVP 1074
```

```
Qy 1171 ALPOTGIPTALVKGSISRMPIE-DS-----SPEKGREEAASKGHVIEYEGKSG----- 1217
Db 1075 GLETTVPGT-SYPPGALTQDEVMDTWTMLMSTIRPTVVSSTSESEVIEYAEAGSSPTFAS 1133
Qy 1218 -----HI-----LSYDNI-----KNAREGTRSPRTAHEISL----- 1243
Db 1134 TLRPFQTHVQLMEETTEEGKASLDVTLGSLGFEPERATELPRKFPPTSPDISVFTAIDS 1193
Qy 1244 -----KRSYESVEGNIKOGMSRSPSPAPLEGLICRALPRGSP 1282
Db 1194 LHRTPPLSPSSFTTEQORVFEESSEKTDGILPGESVQHPVTLTLDIVAMKT----- 1247
Qy 1283 HSDLKERT-----VLSGSIQMGTPRATTEFEDGLKYPKQIK-----RESPP 1324
Db 1248 ESDIDHMTSRPPVTOPTPRPSVVERKTTSKQELSTSPAACTKFPHDINVIIEVRENKT 1307
Qy 1325 IRAFEGAITKGKPYDGIITIKEMGRSIIHEIPRODILTQESRKTPEVVVOSTRPI--IEGSI 1382
Db 1308 GRRLSDMIVSGHPID-----SEKKEEPCSEETDPLDLFAEI 1345
Qy 1383 SQGTPIKF-----DNNSGQSAIKHNKSLITGPS-KLSRCMPPLLEIYENIKVVE- 1431
Db 1346 LPELPDSFEIDIYHSEDEDEGEDCV--NATDVTTPTSVOYINGKQLVTTVPKDEPAAEA 1403
Qy 1432 -RGKYEDVKAGETVRSRHTSVVSSGPSVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYQN 1490
Db 1404 RRGQYESV-----APSQN-----FPDSSATDI---HQF 1428
Qy 1491 TMSGSPMMNKTSDVTTPPNKSTNHRKSTLT--PTQRESIP--AKSPVPGVDVPVWSHP 1546
Db 1429 ILA-----ETESSTTMQFKKSKEGTELLEITWKRPETYPETPDHVSGCEDPVFTLS--- 1479
Qy 1547 FDPHHRGTADEVVWSHLPTQLDPAM--PEHRALDPAAAVLFOROLSPTPGYSQYQ-L 1603
Db 1480 ---SHDKT---TRWSESITESSNLENPVH-----KQKPVLPFFEESSGE 1520
Qy 1604 YAMENTRQ-----TILNDYITSOQMOVNL---RPDVARGLSPREQ 1640
Db 1521 GAIEQASQETILSRATEVALGKETDQSTLTSSILSSSVSVNLEEPLTLTGISQDTE 1580
Qy 1641 PL-----GLPYATRGID-----LTNNMPTTILVPHPGGTSTP- 1673
Db 1581 SMSTIESWEITPSQTVKFSESSAPIIEGSEVEENKNKIFNNMTDLPQRPDPTDLSPL 1640
Qy 1674 PMDRI-----TYIPGTQITFPPRPYNSASMP--GHPTHLAAAAS-----AERERERE 1721
Db 1641 DMSKIMITNHHIYIPAT-----IAPLSKLPSPDARPTTVWNSNSTSEWSDSKSFEGRKK 1695
Qy 1722 KERERE-----RIAAASSD-LYLPRGSEQPCRGCS-----HGYVRSPSPSVR 1762
Db 1696 KENEDEGAVNAAHQGEVRAATERSDHLTLTPELESSNVDASSLATWEGFILETTPT-E 1754
Qy 1763 TOETMLOQRPVSFOGTNGTSVI--TPLD-----PTAQLRIMPLPAGGPSISQGLPASRY 1814
Db 1755 SEKEMANSTP--VFRETIGVANVEAQPFHESSSSHPRQVEELTTLGPNPPLFTDLGSDA 1813
Qy 1815 NTAADALAALVDA-----AASAPQMDVSKTESKHEAARLENLSRSAV 1860
Db 1814 STGMELITASLFTLDLSETKVKKELPSTPSPSVEIS-----SSFEPTGLTPTSTVLDIEIA 1869
Qy 1861 SEQOQLBQKTELEV-----KRSVQCLYTSAPPKPPKQPHSSVWYSEAGKDKGPPPKS 1913
Db 1870 GVMSQTSQKTLISISCKPVSQSGVRDLYT--GPPMGE--DFSGDFSEYTVSYPTMK- 1923
Qy 1914 RYEBELTRGKTTTAAANFIDVITROIASDKDAREGSSQSSSSSLSSHRYPETPSDAI 1973
Db 1924 --EFTVMGG-----SD-DERVRDQTQSSIPTTSSNIYPPVPSKG 1961
Qy 1974 EVTSPASSPAPPOEKLQ-----TYOP-----EV 1996
Db 1962 PDSTVASTAPPWEVEVSSAEGGSQLASVRSSVGPVPLAVDIFSGTESPYFDEEFEEV 2021
```







RL. Nat. Genet. 14:33-41(1996).  
CC - FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.  
CC - SUBCELLULAR LOCATION: NUCLEAR.  
CC DISEASE: PARTICIPATES IN A T(8;16)(P11;P13) CHROMOSOMAL  
CC TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMAERA OBSERVED IN THE  
CC M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.  
CC - SIMILARITY: CONTAINS 2 PHD ZINC-FINGER DOMAINS.  
CC - SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U47742; AAC50662.1; -;  
CC HSP; G060631; 4GBQ.  
CC MIN; 601408; -;  
CC InterPro: IPR001965; -;  
CC InterPro: IPR002717; -;  
CC Pfam: PF01853; MOZ\_SAS; 1.  
CC Pfam: PF00628; PHD; 2.  
CC  
CC Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;  
KW Nuclear protein.  
FT ZN\_FING 206 256 PHD-TYPE 1.  
FT ZN\_FING 259 313 PHD-TYPE 2.  
FT ZN\_FING 371 379 POLY-SER.  
FT ZN\_FING 538 560 C2HC-TYPE.  
FT DOMAIN 788 1478 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 788 798 POLY-GLU.  
FT DOMAIN 989 995 POLY-GLU.  
FT DOMAIN 1019 1026 POLY-ARG.  
FT DOMAIN 1069 1078 POLY-GLU.  
FT DOMAIN 1147 1150 POLY-LYS.  
FT DOMAIN 1225 1228 POLY-GLU.  
FT DOMAIN 1267 1275 POLY-GLU.  
FT DOMAIN 1294 1302 POLY-GLU.  
FT DOMAIN 1411 1414 POLY-GLU.  
FT DOMAIN 1593 1597 POLY-SER.  
FT DOMAIN 1647 1701 GLN/PRO-RICH.  
FT DOMAIN 1647 1650 POLY-GLN.  
FT DOMAIN 1651 1656 POLY-PRO.  
FT DOMAIN 1659 1668 POLY-PRO.  
FT DOMAIN 1668 1675 POLY-PRO.  
FT DOMAIN 1690 1696 POLY-PRO.  
FT DOMAIN 1707 2004 MET-RICH.  
FT SITE 1546 1547 BREAKPOINT FOR TRANSLLOCATION TO FORM  
FT MOZ-CBP.  
SQ SEQUENCE 2004 AA; 225054 MW; 9FFBBAC3792854BA CRC64;

Query Match 2.58; Score 310.5; DB 1; Length 2004;  
Best Local Similarity 17.18; Pred. No. 0.00013;  
Matches 358; Conservative 276; Mismatches 640; Indels 823; Gaps 87;

QY 55 QLLQQQQQLRRPRLSEPHGSDGRPQRTSYEPF---HPGFS-----PVDHD 102  
DB 318 KLLQKAAQIKRRYTNIG--RPNRLKKQNTVSKGPFVKVTPGGRKKKIYLLSQSA 375  
QY 103 SLESKRRLBOVSDSHFORVSAVL-----PLVHPLPEGLRASADAKDPAGF----- 150  
DB 376 SSSSEGYLERIDGLDFCRDSNVSLRNFNKKTKGLDGLTKFTFPSPDGRK--ARGEVDY 433  
QY 151 -----CKHEAPSPISGQPCGDDONASPKLSKEELI-----QSMRDVREIAK-VQOQI 199  
DB 434 SEQYRIKRGNRKSSSTSDWPT-DNQDQWDGKQENEEERLFGSQEIMTEKMLFRDIQEQ 492  
QY 200 LKLLKKQQLLEEAAPKEPEKPV-----SPRPVE----- 229  
DB 493 LQ-----KVGVTGPPDQVRCPSVIEFGKVEIHTWYSSPYQEYSLPKLYLCEFL 544

QY 230 --QKHSIVQIYIDENRRKKAEEAHKIFGLGPKVPELYNQPSD--TKVYHENI----- 279  
DB 545 KYMSRTILO---QHMKKCGWFPPANEIYRKNNIISVFVDGNGVSTIYCNLCILAKLF 600  
QY 280 ---KT-----NOVMRKKLILFFKRRNHARKOREQKICQRYD---OLMEA 317  
DB 601 LDHKTLYYDVEPFLFYVLTQNDVKGCHLVGVFSKEKHQ---QKYNVSCIMILPQ 652  
QY 318 WEKK---VDRIENPRRAKESKTYEYKQFPEIRKQREOQERFORVQORCA--- 367  
DB 553 YORXGYGRFLID-----FSYLLSKRE-----GQAGSPEPK 682  
QY 368 ---GLSATIARSEHSEIIDGLSEQENNEKQRLSVIPPMFMFDAQRRVKFTNMNGL 423  
DB 683 LSDGLRLSYMAYWKSVILECLYHQNDKQISIKLSKLTGICPDQITSLHLHRLMD--- 738  
QY 424 MEDPMKYVKDRQPMNVWTDHEKEIFKQFIQHPKNFGLIASYLERKSV-PCVLYYYITK 482  
DB 739 -----FRSDQFV-----IIRREKLIQ--DHMAKLIQLNLRPVDVDPCLRWTPVIV 781  
QY 483 KNEYKALVRRNYKRGGRNQOIARPSQEEKVEEKEADAEKTEKKEEKDEE----- 536  
DB 782 SN-----SVVSEEEEEEEEGEENEEPPQOERELEISVG 814  
QY 537 -----EKDEKEDSKENTKEKDIDGTAEETEERQATPRGRKTANSQGRRKGRITR--- 587  
DB 815 KSVSHENKEQDSYSVESEKKPEVMAPVSSTRLKSQVLPHDSLANSQPSRRGRWGRNKR 874  
QY 588 -----SMTNEAAAAASAAAAATEPPP-----PP 610  
DB 875 TQERFGDKSKILLEETSSAPQEQYEGEGEKSEATQEQYTESEQLVASEEQSDGKPD 934  
QY 611 LP-----PPPEPISTEPVETS---RWTEEMEVAKKGLVEHG----- 644  
DB 935 LPKRLSEGVEPWGQLKKSPALKCRLTGSERLPRYSEGDRAYL-RGSESESEEEE 993  
QY 645 -RNWAAIAKVMGTKEAOCKNFYFNKRR-----HN-----LDNLL----- 679  
DB 994 PESPRSSPPILTKPTLKRKKPFLHRRRRVRKRKHNSVYVTTTSETTEVLDEPFEDSD 1053  
QY 680 -----QOHKOKTSRKPRE-----ERDVSQESVASTVSAQEDDETEA 716  
DB 1054 SERPMPRLPTFEIDEEEEEENELPPREYFRRLSSQDVLRCQSSSKRKSDEESES 1113  
QY 717 SNEEENPEDSEVEA-----VKPSEDSPENAT-----SRGTEPAV 751  
DB 1114 DDADDTPLIKPVSLLRKRDVKNSPLEPDTSTPLKKGKWKPKSRKPIHWKKRCKKPGF 1173  
QY 752 ---ELEPTTTAFSTSPSLAVST--KP---ADESVETQVNDNISAEATAEQMDVDQOE 802  
DB 1174 KLSREIMPVSTQACVIEPIVSIKAGRKPKIQESEETVEPKEDMPLPERKEEEMQAEA 1233.  
QY 803 HSAEGSVCD-----PPATKADSVDFVVRYPENHASKVEGDNTKERDLDRASEKVE 854  
DB 1234 EEAEEGEEEDAASVEVPAASPADSNSPETETKEPE-----VE-----EEEEKRVSEQR 1284  
QY 855 PRDEDLVVAQINAQRPPEPQSDNDSSATCSADEVDGSEPERQRMFPMDSKPSLLNPTGSI 914  
DB 1285 QSEE-----EQELEPEPEEEDAAATAQNDHDDHDADEDDGHLESTKKE----- 1331  
QY 915 LVSSPLKNPLDLPOLQHRAAVIPPVMVSCPTCNPICPTVPSGYALYQRIKAMHESALLE 974  
DB 1332 LEEQPTREDVKKEEPGQVE-----SFLDANMQS-----REKIDKEETEELDS 1373  
QY 975 EQRQREQIDLECRSSTSPCGTSKSPNREWEVLQAPAPHLITNLPBGVRLTPTTRPPPP 1034  
DB 1374 EEEQ----- 1377  
QY 1035 PLIPSSKTTVASEKPSFTMGGSISQGTPTGTYLTSHNOASYTOETPKPSVGSISLGLPRQQ 1094  
DB 1378 ---PSHTSVVSEQ---MAGSED-----DHEEDSHYKE----- 1404  
QY 1095 ESAKSATLPYIKOEFSRPNQSONSQPEGLLVRAQHEGVVVRGTAGATQEGSITRGTPTSKIS 1154



Db 468 SJSN---PNPIDPSSMORAYAAALGLPYMNOPTQLOPQVPGQQAQPPAHO---QMRTLN 521  
Qy 422 GLMBDPMKV-----YKROFMNVWTDHEKEI---FKBKFIQHPKNFGLIASYLERKSVPD 473  
Db 522 ALGNPNMSPVAGGITTOQPPNLSIALPTSLGATNPLMDGNSNGISL---STIP- 577  
Qy 474 CVLYYYLTKKNENYKALVRRNYGRRKRNOQIARPSQEKVEEKEE-----DKAEKTEK 527  
Db 578 -----TAAPPSSGTVRKWHEHVTDQLRSHLVHKLVAQIFPTTDPDAALKDR 623  
Qy 528 KEEE-----KKDEEKEDEKEDSKENTKE--KDKIDGTAEEETEEREQATPRGRKTANQGR 580  
Db 624 RMENLVAAYAKVEGDMYESANDEYHLLAEKIKYTKOLEEKR-----RTLHKQG- 676  
Qy 581 RKGRITRSMTEAAAAAATAEPP-----PPLPPPPPIETEPVETSRWTEEMEVEV 635  
Db 677 -----ILGNQFALPASGAQPPVIPPAAQSVRPNGLPLPLPVNRMVQSGMNSFNMSL 728  
Qy 636 AKGLVHGRNMAAIAM---VGTKEBAQCKNFYNTKRRHNLNLLQHKQK--TSRKPR 691  
Db 729 GNVQLPQAPMGPRASPMNHSVQMSNASVPGMAISPSRMPQPPNMMGTHANNIMAQAPT 788  
Qy 692 EERDVSCQ-----SVASTVSAQEDIEDIASNEBE-----NPEDSEVAV 731  
Db 789 QNOFLQNOFPSSSGAMSVNSVGMGPAQAQAGVSGQGPAAALPNPLNMLAPOASOLPC- 847  
Qy 732 KPSEDSP-----ENATSRGNT--EPAVELEPTTETAPSTSPSLAVPST 772  
Db 848 PVTQSPHLPTPPASTAAGMPSLQHTAFTGNTPPQAPATQPTPVSSGOTPT-PTPGS 906  
Qy 773 KPAEDESVEQVNDISIAETAQMDVDQOEHSAEESVCDP-----PPAT-- 817  
Db 907 VPS---AAQTOSTPTVQAAQAQV--TPQPTPVQPSVATPQSSQOQPTPVHTQPPGTP 962  
Qy 818 ---KADSDVDVVRVENASKVEGDNTERDLDRASEKVEPRDEDLVVAQINAPRPPQS 875  
Db 963 SOAASIDNRVPTSTVTSATSSQOPGPDVPMLEMTQVOTD-----AEPEPTE 1013  
Qy 876 DNDSSATCSADEVDGPEPORMPMDSKPSLLNPTGSIILVSSPLKNPLDLPOLQRAA 935  
Db 1014 SKGPRSEMEEELQGSQVKEETD-----TTEQKSEPMVEEKKPEVK 1057  
Qy 936 VIPMWSCPTCNIPDIPVSGYALYQRIKAMHESALLEQROREQIDLECRS--STSPC 994  
Db 1058 V-----EAKEEENSNDTASQSTSPS 1079  
Qy 995 GTSK---SPNREWEVLOPAPHOLLITNLPQVRLPTTRPPPLIPSSKTTTVAERKPSF 1051  
Db 1080 QPRKKIFKPEELQALMPTLEALYRQDPES--LPFRQPVDPQLLGIPIDYDIVKNPMDLS 1137  
Qy 1052 IMGGSISOG---TPGTV-----LTSNHOASVTOETPK-----PSVGS 1086  
Db 1138 TIKKRLTQGOEYQWQYVDDVRLMNFNAWLNRYKTSRVYKFCSLAEVFEQEIQVPMQSL 1197  
Qy 1087 SLGLPROQESAKSATLPIKOEFSRPNQSQEGLLVRAQHEGVRGTA--GAIQESIT 1145  
Db 1198 GYCCGRKVEFSQTLCCYKQLCTIPR-----DAAYSVYQNRVHFCGKCFTEIQENVT 1251  
Qy 1146 RGTPTSKISVESIPSLRGSITQGTTPALPOTGIPTALVKGSIRMPDIEDSSPEKGREAA 1205  
Db 1252 LG-----DDPSQPTTISKDQFEKKNDTL--DPEPPFVDCRECG 1288  
Qy 1206 SKGHVIEGKSGHILSYDNI-----KNAREGTRSPRTAHEISIKRSYESVEGN----- 1253  
Db 1289 RKMHCIC-----VLHYDITWPSGFVCDNCLKTKGRPRKENKFSAKRLQTLRLGNHLEDR 1342  
Qy 1254 IKOGMSRSPVSAPEGLICRALPRGSPHSD-----LKERTVLGSGSIMOGTTPRATT 1305  
Db 1343 VNKFLRQNHPEAG-----EVFVRVASSDKTVEVKPKMSRFVDSGEMSESFYRTK 1395  
Qy 1306 ESFEDGLKYPKQIKRESPPTRAFEGAITKPKPYDGIITIKEMGRSINE-----IPQDILT 1361  
Db 1396 ALF-----AFD-----EIDGV--DVCFPGMHVQDITALIAPHQ----- 1425

Qy 1362 QESRKTPEVQOSTRPIIEGS---ISQGTPIKFDN--NSGQSAIKHN-----VKSIL- 1407  
Db 1426 -----TOGCYVYISYLDISIHFRPRCLRTAYVHEILIGLYLEVVKLVYVT 1469  
Qy 1408 -----TGPSKLSRGMPPLEIVPENIKVVERGK-----YEDV--KAGE 1442  
Db 1470 AHIWACPPSEGDDYIFHCHPPDPQIKPKRQEWYKKMLDKAFARIINDYKIDIFKQANE 1529  
Qy 1443 TVRSRHTSVV-----SSGPSVLRSILHEAPKAQLSPGIYDDTSARETPVSQO---NT 1491  
Db 1530 ---DRLTSAKELPYFEGDFWPNVLEESIKELEQEEERKKEESTAASETPEGSOGDSKNA 1586  
Qy 1492 MSRGSPMMNRT--SDVT-----IPPNKSTNHERKSTLTPQRES-----IPA 1531  
Db 1587 KKKNNKTKNKKSSISRANKKPSMPNVNDSQKLYATMEKHKEVFFVIHLHAGPVIST 1646  
Qy 1532 KSPVPGVDVPPVS-----HSPF-----DPHHRGSTAGEVVWS----- 1562  
Db 1647 QPPIVDPDPLLSCLMDGRDAFLTLARDKHWFSLSRRSKWSTLCMLVELHTQGDREYV 1706  
Qy 1563 -----HLPTQ-----LDPAWPFHAL-----DPAAYL 1586  
Db 1707 TCNECKHHVETRWHTVCEYDLCINCYNTKSHTHKMVKWGLGLDDEGSSGQSPQ 1766  
Qy 1587 FORQLS-----PTPGYPSQYLYAM 1606  
Db 1767 ESRLSITORCISLVHACOCRNANCSLPSCQKMKRVVOHTKCKRKTNGCGPCVKQLIAL 1826  
Qy 1607 -----ENTROTILNDYITTSQMQ-----VNLRPDVARGL--SP 1637  
Db 1827 CCYHAKHCQENKCPVPFCLNFKHNVRQOQIQHCLQAQALMRRMATMTRNVPOQSLSP 1886  
Qy 1638 REQPLGLPYPATRGIDLTNMPPILYVHPHGGTSTP--PMDRI--TYIPGTQITFPB- 1690  
Db 1887 TSAPPGTP-----TQOPSTPQPPPAQPPSPVNNMSPAGFPNVAQTPTTIVISA 1936  
Qy 1691 -RPNYSASMSPGHPTHLAAASAEERERERERERERERERERERERERERERERERERER 1749  
Db 1937 GKPTNQVPAPPPAPPPAAVEAARQIEREAQOQHLYR---ANINNGMPGCRDGMGTGP 1993  
Qy 1750 SHGVVRSPSPSVRTOEMLQORPSVFGTNGTSTVITPLDPTAQLRIMPLPAGGSPISQGL 1809  
Db 1994 SQ-----MTPVGLNVRPNQVSGPVMSM-----PPCQMQQAPIQOQP--MPGM 2036  
Qy 1810 PASRYNTAADALAAALVDAASAPQMDYSKTESKHEARLEENLRSAAVSEQOOLQOK 1869  
Db 2037 P-----RPVMSMOQAQAAVAGPRMPNVQPNRSISPSA-LQDLTLTKSPSPQOQ--QQ 2086  
Qy 1870 TLEVEKRSVQCLYTSSAF-----PSGKQPQ--H-----SSV 1898  
Db 2087 VLNILKSNPQLM---AAFIKORTAKYVANQPMQPOGLOQSQPMQPOGPMHQPSLQNL 2143  
Qy 1899 VYSAQKDK-GPPPKRSYEEELRTGKTTITPAANFID-----VIITROI- 1941  
Db 2144 NAMQAGVPRPGVPPQPMAGGLNFGQO---ALNIMNPGHNPNTNMNPQYREMYRROLL 2199  
Qy 1942 --ASDKDAREGSSOSSSSSLS--HRYETPSDAIEVISPAS--SPAPPOEKLQYQ 1993  
Db 2200 QHOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2253  
Qy 1994 P-----EVVYKANO---AENDPTRO---YEGPLHHRPOOESPSPOOQLPP 2032  
Db 2254 PIQSSMGQMAAPMQLGQMGQPGGLGADSTNIOALQOORILQOQOQOQOQOQOQOQOQO 2313  
Qy 2033 SSOAEGMGQVPRHRLITLADHI--CQIITQDFARNQVSSQTPQOQPTSTFQNSPSALVS 2090  
Db 2314 SPQOQHLSGQOQO-----ASHLPQOQIATS--LSNQVRSAPVQSPRPQ-----S 2355  
Qy 2091 TPVRTKTSNRYSPESQASQSVHH-----QRPGRSVSPENLVKSRGSRPGKS----- 2136  
Db 2356 QPPHSSPSPIQOQO---PSPHHVSPQTGTPHGLAVTMASMDQGHGHNPEQOASMLPOLN 2412













CC . CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS  
CC AT A DEFINED SITE, PCE/THE, WITHIN THE HCF REPEAT (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D45419; BAA08258.1; -  
CC HSSP; P02751; 1FNA.  
CC InterPro; IPR001777; -  
CC Pfam; PF00041; fn3; 2.  
CC Nuclear protein; Repeat.  
CC FT REPEAT 44 89 KELCH 1.  
CC FT REPEAT 93 140 KELCH 2.  
CC FT REPEAT 148 194 KELCH 3.  
CC FT REPEAT 217 265 KELCH 4.  
CC FT REPEAT 266 313 KELCH 5.  
CC FT DOMAIN 1010 1448 8 X 26 AA APPROXIMATE REPEATS.  
CC FT REPEAT 1010 1035 HCF REPEAT 1.  
CC FT REPEAT 1072 1097 HCF REPEAT 2.  
CC FT REPEAT 1101 1126 HCF REPEAT 3.  
CC FT REPEAT 1157 1182 HCF REPEAT 4 (DEGENERATE).  
CC FT REPEAT 1295 1320 HCF REPEAT 5.  
CC FT REPEAT 1323 1348 HCF REPEAT 6.  
CC FT REPEAT 1358 1383 HCF REPEAT 7 (DEGENERATE).  
CC FT REPEAT 1423 1448 HCF REPEAT 8.  
CC SEQUENCE 2090 AA; 214942 MW; E495E8B1F2385E17 CRC64;  
  
Query Match 2.4%; Score 300.5; DB 1; Length 2090;  
Best Local Similarity 19.5%; Pred. No. 0.00033;  
Matches 338; Conservative 193; Mismatches 641; Indels 565; Gaps 69;  
  
QY 966 AMHESALLEQRORQEOIDLECRSSTSPCGT-----SKSPNREWEVLQAPAPQL 1014  
DB 388 ATADSYLLQKY---DIPATATATSTPNPVPSPANPKSPAPAAAAPAVQPLTVQG 444  
QY 1015 ITNLPEGVRLPTTRPPPLPSSKTV-----ASEKPSF--IMGGSISQGTPTLT 1067  
DB 445 ITLVQAAAAPPSTTTIQLVTPVGSISVPTAARAGVPAVLKVTGQATTGTPLTMR 504  
QY 1068 SHNOASTQETPKPSVSGISLGLPRQESAKSATLPYKOEESPRQNSQOPEGLLVRAQ 1127  
DB 505 PAGQAGKAPVTVTSLPASVRMVVPTQ--SAQGTVI-----GSPNQMSGMAALA- 550  
QY 1128 HEGVVRGTAGAIQSGSITRTPTSKISVESTPS-----LRGSIQTGPALPOTGIPTAL 1182  
DB 551 -----AAAAATQ-----KIPPSAPTVLSPAGTTIVKVVAVTPGTTLPAT----- 592  
QY 1183 VKGISRMPEDSSPERGEEAASGHVIEGKSHILSYDNINKNAREGTRSPRTAHE-- 1240  
DB 593 KVVASSPVVSNPATRLMKTAQAQVGSVSSA-----ANTSTRPIITVHKSG 639  
QY 1241 -ISLKRSE---SVEGNKMGMSRESVPSAPLEGLICRALPRGSPHSDLKERTVLSGSI 1296  
DB 640 TTVTAQAQVVTTVVGGVTKTITLVKSPISVPGGSALISNL--GKVMVSVQTKPVQTSV 697  
QY 1297 MQGTPRATTSFEDGLKYPKIKRESPIPAFEGAIK-----GKPYDGTITIKGRS 1350  
DB 698 ---TQASTGPV-----TQIOTKPLPA--GTILKLVTSADGKPTTIITTTQASGAG 745  
QY 1351 -----IHEIPRODILTQESR-----KTPGVOSTRPII 1378  
DB 746 SKPTILGISSVSPSTTKPGTTTIKTIPMSAITIQAATGVTSTPGIKSPIITITTKVMT 805  
QY 1379 EGSISQGTPIKFDNNSSQSAIKHNKVSILITGPKSLRGMP--LEIIVP-ENIKVVERGKY 1435  
DB 1379 APADSLNDPSTESCLN-----ELASAVPST-VALLPSTAT 1807

DB 806 SGT---CAPAKIITAVPKIATGHGQGVTVQVVLKAGPQPGAILRTPVMSGVRLVTPVT 862  
QY 1436 EDKAGETVRSRTSVV--SSGPSVLSTLHEAPKAQLSPGIYDDTSARRTPVSVQNTMS 1493  
DB 863 SAVKPAVT---TLVVGKGTGTTGLTVGTSTSLAGAHSTSLATPITLITGLTIA 917  
QY 1494 RGSFMMNRTSDVTIPPNNKSTNHERKSTLT-----PTQRESIPAKSPV- 1535  
DB 918 -----TLSSQVINPATAIVSAOATTLTAAGGLTTPITMQPVSQPTQVTLITAPSGVE 970  
QY 1536 -----PC-VDPVWSHSPDPHRRSTAGEV 1559  
DB 971 AQPVHDLPVLSILASPTTEQPTATVTIADSGQGVQDQVTVTLVCSNPPCETHTGTT 1026  
QY 1560 YWSHLPTQLDPAMPFHRALDPAAAAAYLFQRLSPTPGPSQYQLYAMENRTQITLNDYIT 1619  
DB 1027 -----NTATTTVANLGHP-OPTOVQFVC---DRQEAASLVT 1061  
QY 1620 SQMQVNLDPDVARGLS--PRE-OPLGLPYPATRGIIDLTM-----PPTILVPHPGGT- 1670  
DB 1062 SAVGQON--GNVVRVCSNPPCETHTGTTTATATATSNMAGOHGCSNPP--CETHETGTT 1117  
QY 1671 -----STPPMDRITYIPG---TQITPPRPYNASMSPGHPT 1704  
DB 1118 STATTAMSGTGOORDTRHTSSNPTVVRTVAPGALERTQGTVPKQ---COTQOANMT 1173  
QY 1705 HLAASAAERERERERERERERERERERERERERERERERERERERERERERERERER 1764  
DB 1174 NNTMTVQATR-----SPCAGPALLRPSVALEAGNHSFAFVQLALPSVRV- 1217  
QY 1765 ETMLQRPVSFQGTNGTS-----VITPLDPTAQLRIMPLPAGGPSISQGLP 1810  
DB 1218 -----GLSGPSNKMPTGHOLETVHTVTNTPTTALSIM--GAGELGTARLIP 1263  
QY 1811 ASRYNT-----AADALAAALVDAASAPQMDVSKTKESKHEAARLENLRSRAV 1860  
DB 1264 TSTYESLOASSPSTMTWTALEALLCPSATVTVQVCSNPPCET-HETG--TTNTATTSNAG 1320  
QY 1861 SEQOOLEKTLVEKRSVQCLYTSSAFPSGPKOPHSSVVYSEAGKD----- 1906  
DB 1321 SAQVCSNPPCETHTGTTTATATATSNAGAQV-----EGQQPAGGRPCETHOTS 1373  
QY 1907 -----KGPPKSRYEELTRGKTITA---ANFIDVITRIADSK---DA 1947  
DB 1374 TGTMSVSVGALLDPATPSHCTLESGLVAVSTVSOAGATILLASFTQVCSNPPCET 1433  
QY 1948 RERGS---QSSDSSSLSHRYETP--SDAIEVIS----- 1977  
DB 1434 HETGTTTATTTVSNMSSNODPPPAASDQGEVSTVQGSANITSSSGITTTVSTLPRAV 1493  
QY 1978 -----PASSPAPPQEKLOTYQP-----EVVKANQANDPTROYEGPLHY 2017  
DB 1494 TTVTQSTPVPGPSVPNTSSLTETGALTSEVPITATITVTANTTETSDMPFSAVDILQPP 1553  
QY 2018 RPOQESPOOOLPP-----SSQAEGMQVPRHRLITLADHICQIITODF 2063  
DB 1554 EELQVSPGPRQLPQLQLQSSATPLMGESSEVLASQTPELQAAVDL----- 1601  
QY 2064 ARNOVSQTPOQPTSTFQNSPSALVT-----PVRTKTSNRYSP-SSQASVHHQ 2113  
DB 1602 --SSTGPFSSGQETS-----SAVATVVVQPPPTQSEVDQLSLPOELMAEAQAGTTT 1653  
QY 2114 RPSGRVSPENL-----SAVATVVVQPPPTQSEVDQLSLPOELMAEAQAGTTT 2132  
DB 1654 LMVTLPEELAVTAAAEAAQAAATTAQVLAQAAQVMACTGPMPTSEAAA 1713  
QY 2133 PGKSPERSHVSSEYEPISPPQVFWHEKQDLSLLLSQ-----GAPBQBN---DAR 2183  
DB 1714 AVTQAEGLHLSAQEGE-QATTIPVLVQAEALVQOQOQLQEVQAAQOQHHLPTAL 1772  
QY 2184 SPGISYLPSTFTKLENTSPVMSKKQBIKRLNSSGGSDMAAAQPGTIFNLPAVTT 2243  
DB 1773 APADSLNDPSTESCLN-----ELASAVPST-VALLPSTAT 1807







DR EMBL; D17635; BAA04538.1; JOINED.  
 DR EMBL; D17636; BAA04538.1; JOINED.  
 DR EMBL; D17637; BAA04538.1; JOINED.  
 DR EMBL; D17639; BAA04538.1; JOINED.  
 DR EMBL; D17640; BAA04538.1; JOINED.  
 DR EMBL; D17641; BAA04538.1; JOINED.  
 DR EMBL; D17642; BAA04538.1; JOINED.  
 DR EMBL; D17643; BAA04538.1; JOINED.  
 DR EMBL; D17644; BAA04538.1; JOINED.  
 DR EMBL; D17645; BAA04538.1; JOINED.  
 DR EMBL; D17646; BAA04538.1; JOINED.  
 DR EMBL; D17647; BAA04538.1; JOINED.  
 DR EMBL; D17648; BAA04540.1; JOINED.  
 DR EMBL; D17634; BAA04540.1; JOINED.  
 DR EMBL; D17636; BAA04540.1; JOINED.  
 DR EMBL; D17637; BAA04540.1; JOINED.  
 DR EMBL; D17639; BAA04540.1; JOINED.  
 DR EMBL; D17640; BAA04540.1; JOINED.  
 DR EMBL; D17641; BAA04540.1; JOINED.  
 DR EMBL; D17642; BAA04540.1; JOINED.  
 DR EMBL; D17643; BAA04540.1; JOINED.  
 DR EMBL; D17644; BAA04540.1; JOINED.  
 DR EMBL; D17645; BAA04540.1; JOINED.  
 DR EMBL; D17646; BAA04540.1; JOINED.  
 DR EMBL; D17647; BAA04540.1; JOINED.  
 DR EMBL; D17552; BAA04490.1; JOINED.  
 DR EMBL; M26684; BAA04811.1; JOINED.  
 DR PIR; A32642; A32642.  
 DR HSP; P05412; IFOS.  
 DR InterPro; IPR000075; -.  
 DR Pfam; PF02029; CALDESMON.  
 DR PRINTS; PR01076; CALDESMON.  
 KW Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;  
 KW Repeat; Alternative splicing.  
 FT DOMAIN 26 199 MYOSIN AND CALMODULIN-BINDING.  
 FT REPEAT 251 390 10 X 13 AA APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 251 265 1.  
 FT REPEAT 266 278 2.  
 FT REPEAT 279 291 3.  
 FT REPEAT 294 306 4.  
 FT REPEAT 309 321 5.  
 FT REPEAT 324 336 6.  
 FT REPEAT 337 349 7.  
 FT REPEAT 350 362 8.  
 FT REPEAT 363 375 9.  
 FT REPEAT 378 390 10.  
 FT DOMAIN 523 580 TROPOMYOSIN-BINDING (POTENTIAL).  
 FT DOMAIN 622 632 TROPOMYOSIN-BINDING (POTENTIAL).  
 FT DOMAIN 612 644 STRONG ACTIN-BINDING.  
 FT DOMAIN 674 680 CALMODULIN-BINDING.  
 FT DOMAIN 726 752 WEAK ACTIN-BINDING.  
 FT DOMAIN 39 46 POLY-ARG.  
 FT DOMAIN 539 542 POLY-GLU.  
 FT DOMAIN 556 559 POLY-GLU.  
 FT MOD\_RES 485 485 PHOSPHORYLATION (POTENTIAL).  
 FT MOD\_RES 597 597 PHOSPHORYLATION (BY CDC2).  
 FT MOD\_RES 682 682 PHOSPHORYLATION (BY CDC2).  
 FT MOD\_RES 688 688 PHOSPHORYLATION (BY CDC2).  
 FT MOD\_RES 711 711 PHOSPHORYLATION (BY CDC2).  
 FT MOD\_RES 717 717 PHOSPHORYLATION (BY CDC2).  
 FT VARSPIC 1 24 MDDFERRELRRRQREEMLEAER -> MISRSYCRQNL

Query Match 2.4%; Score 298.5; DB 1; Length 771;  
 Best Local Similarity 20.9%; Pred. No. 0.00012;  
 Matches 182; Conservative 124; Mismatches 334; Indels 231; Gaps 34;  
 QY 293 FRRNHARKQREKICQRYDQLMEAWKKVDRIE---NNPRKAKESKTR-----E 340  
 DB 4 FERRELRRRQREEMLEAERL--SYQRNDDDEEAARRRRRRRQRRRQREKEDGVSE 61  
 QY 341 YFEKQFPEIRKQREOQRFQVQRGAGLSATIAARSEHSEIIDLSEQENNEKQMRQL 400  
 FT

Db 62 VTEKSEVNAQNSVAEEETKRSTDDEAALLERLARREERRRQKRLQEALEKQKEDPTITDG 121  
 QY 401 SVIPPMFDAEQRRVKFINMGL-----MEDPMKYVKDQFMNWTDEH- 444  
 Db 122 SLSP-----SRRENNVEENITGKEEKVETROGCEIEETETVTKSQRNWRODGE 176  
 QY 445 --KEIFKDKFIQHPKNFGLIASYLERKSVDCVLYYYLTKKKNYKALVRRNYGKRGRN 502  
 Db 177 EGKKEEKDEEEKPEVP-----TEENQVDVAV-----EKSTDKSEEVVETTKTLAVNAEN 225  
 QY 503 -----QIARPSQEEKVE-EKEEDKAEKTEK-----EKEKDEEKEDEKESKENT 548  
 Db 226 DTNMLEGEQISITDAADKEEAEKEREKLEAEERKAAEEKAAEEKAAEEKAA 285  
 QY 549 KKKDKIDGTAEE--TEERQA--TPGRKKTANSQGRKGRITRMTNEAAAAASAAAAATE 605  
 Db 286 EERERAKAEEERAKAEEERAKAEEERAKAEEERAKAEEERAKAEEERAKAEE 345  
 QY 606 EPPPLPPPEPISTEPVETSRWTEEMEVAKGLVHGHRNMAAIAKMYGTKSEAQCKNF 665  
 Db 346 R-----AKAEEERAKAEEERAKAEEERAKAEEERAKAEEERAKAEE 380  
 QY 666 YFNKRRHNLNLLQHKOKTSRKPREEDVQCESVASTVSAQEDIEDA---SNEEN 722  
 Db 381 ----KRAEEKARLEAKLKEKKKMEKK--AQEEKAQANLLRKQEDKAEKVEAKKESL 434  
 QY 723 PE----DSEVAVKPSEDS-----PENATSRGNTPEPAVELEPTETA 760  
 Db 435 PEKLOPSTKKQVKQKONKQEKAPKEEMKSVDRKRGVPEQKAQNGERE----- 482  
 QY 761 PSTSPSLAVPSTKPAEDSVETQVNDISIAETAQMDVQDQHSABEGSVCDPPATKAD 820  
 Db 483 -----LITPKLKSTENAFGRNLKGAANAAGSEKLEKQEAIVE----- 523  
 QY 821 SYDVEVRYPENIAKVEGNTKRDLDRAKVEPRDEDLVVAQINAORPEQSDNDSS 880  
 Db 524 -LD-ELKKRRERRRILEEEQKKQEEAEERKIREBEKKRMKEETERRAE-----A 574  
 QY 881 ATCSADEVDGPEQRMPMDSKPSLLNPTGSL-----VSSPLKP----- 922  
 Db 575 AEKROKVPEDGVSEKKPF-----KCFSPKGSSLUKEERAEFLNKSQKSKPAHTTA 628  
 QY 923 --NPLDLPOLHRAAVIPPMVS--CTPCNIPITGTPVSGYALYQRIKAMHESALLEEQRQ 978  
 Db 629 VVSKIDSLRLEQYTSVAVGNKAAKPAKPAASDLVPVAGV---RNIKSMWEG----- 677  
 QY 979 ROEQIDLECRSTSPCGTSKSPNREWEVLQAP-----HOLIINLPGVRLPTRPT--RP 1032  
 Db 678 -----NVFSPGGTG--TPNKETAGLVGVSSRINEWLTKTPEGNKSAPKPSDLRP 727  
 QY 1033 -----PPPLIPSSKTTVASEK 1048  
 Db 728 GDVSGKRNLEWQSVSEKPAASSSKVTATGKK 758

RESULT 37  
 SC16\_YEAST  
 ID SC16\_YEAST STANDARD; PRT; 2194 AA.  
 AC P48415;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE MULTIDOMAIN VESICLE COAT PROTEIN.  
 CN SEC16 OR YPL085W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID:4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96017704; PubMed=7593161;  
 RA Espenshade P., Glimeno R.E., Holzmacher E., Teung P., Kaiser C.A.;  
 RT "Yeast SEC16 gene encodes a multidomain vesicle coat protein that

interacts with Sec23p.";  
J. Cell Biol. 131:311-324(1995).  
- FUNCTION: INVOLVED IN THE BUDDING OF TRANSPORT VESICLE FROM THE  
ENDOPLASMIC RETICULUM. THE C-TERMINAL INTERACTS WITH SEC23 AND  
WITH THE CYTOSOLIC DOMAIN OF SCD4. COULD THEREFORE BE A  
CONSTITUENT OF COPII VESICLE COAT. N-TERMINAL OVEREXPRESSION  
CAUSES A LETHAL SECRETION DEFECT.  
- SUBCELLULAR LOCATION: ON THE ENDOPLASMIC RETICULUM AND ON VESICLES  
WHICH BUD FROM IT.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U23819; AAC49088.1; .  
DR SGD; S0006006; SEC16.  
KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum.  
FT MUTAGEN 1058 L->S: IN SEC16-4; TS ACCUMULATION OF ER  
MEMBRANES.  
FT MUTAGEN 1083 L->P: IN SEC16-3; TS ACCUMULATION OF ER  
MEMBRANES.  
FT MUTAGEN 1088 L->P: IN SEC16-2; TS ACCUMULATION OF ER  
MEMBRANES.  
FT MUTAGEN 1230 W->R: IN SEC16-1; TS ACCUMULATION OF ER  
MEMBRANES.  
SQ SEQUENCE 2194 AA; 241613 MW; BB1E02D2AD4683E3 CRC64;  
-----  
Query Match 2.4%; Score 298; DB 1; Length 2194;  
Best Local Similarity 18.7%; Pred. No. 0.00043;  
Matches 474; Conservative 324; Mismatches 897; Indels 842; Gaps 121;  
QY 14 STEQGRYPHVSQYTFPTTRHOEEFAVDYRSSHLEVSOASOL-LQQOQQOQLRRRPSLL 72  
DB 39 STINSFNDSDSVNRTESDIASKSD--VPPV-SSSTNISPANETQLEIPDTQELHHK--LL 93  
QY 73 SEFHGSDRQRRYSYEPFHGPGSPVDHDS-----LESKRPRLEQ 113  
DB 94 -----NDSQDHITADNSNDLPSNIVEHDSVITOTKPMASOEYEETAHLSSRNPSLDV 147  
QY 114 V-----SDSHFORVSAALPLVHLPLPECLRASADAKDPAP---GGKHEA----- 155  
DB 148 VAGELHNNHETQKTAVS-----AVEEDSFNEEGEENHDSIIITSLND 190  
QY 156 -----PS-----SPISGQPCG-----DDQ-----NASP 173  
DB 191 ATPSQYNHFLPSDGNLLSPELSSGDTPTNNVPLGTDKNEINDDEYCNDEKISLNANNVLP 250  
QY 174 SKLSKEELIQSDRVDRREIAKVEQOILKKKQOOLEEAAKPPPEKVPSPVVEQKHR 233  
DB 251 DELSKEE-----DERLKEITHVSTEERKQDIADOETA---ENLFTSSTEPSSENKIR 298  
QY 234 SIQVLIYDENKKAEEAHKIEGLGPKVELPLYNQPSTKV-YHENIKTNQVMRKLLIF 292  
DB 299 N-----SGDDTSMLFQ-----DDESOKVPEEDVKKD----- 326  
QY 293 FKRRNHARKQREOKTCORYDQLMFAWEKKVDK--IENNRKKAKEKT-----REYEQK 345  
DB 327 FHNETNTNQ-----ESAPNTDORDKGYEGNEALKKSECTAADERSYSEET 373  
QY 346 PPEIKRQREQ-----ERFORVQR--GAGLSATIAARSEHEISEIDGLEQENN 393  
DB 374 SEDIFPHGDKQVGEQNDFTGKNINIENESQKLMGEGNHKLPLSAEADIIE--PGKDIOQA 431  
QY 394 EKQMRQLS-----VTPPMFD-----AEORRVKFINMGLMED-PMKYTKDRQFMNV 439  
DB 432 EDLFTQSGDGLGEVLPWSTOKNADVTSKQSKHEDLFAASGNDKLPWEV-SGGEVSSG 490  
QY 440 WTDHEKEIFDKFIQHPKNF-----GLIASYLERKSVPCVLYLYLT-----K 482

DB 491 KTENSMQSTEKIAQKFSFLENDDDLLDDDSFLASSEEDTVPTDNTNLTSPVEEK 550  
QY 483 KNEYKALVRRNYKRRGR-----NQOIARPSOEKVEE 516  
DB 551 KASRYKPIFEAGMRQEQVHFTNTGIVTPQOFHGLTKTGLGTPNQOVSVNIVSPKPP 610  
QY 517 KEEDKAEKTEKEEKKB-----EKDEKEDSKENTKEKDKIDGTAEETEEQQA 567  
DB 611 VYKDRSNFKINEEKKSDAYDFPLEIISSESKKGHAKPVAVPTQRF-GSGNSFSLLDKP 669  
QY 568 TPRGRKTANSQ-----GRKGRITRSMTEAAAAA-----ATEEP 607  
DB 670 IPOSKSGNSNRPPVPLGTQEPSSR--TNSAISQSPVYAFNPYKIQLOQAQIS 727  
QY 608 PPPLP-----PPP-----EPISTEPVETSRWTEEMEVAKGLVHEGRNMAAIKAMVGTGS 658  
DB 728 GMPLENTNIPPPALKVETTVSAPPTR-----ARGVSNAS 761  
QY 659 EAQCKNFYNTKRRHLDN---LLOQHKOKTSRKPREEDYSOCESVASTVSAQEDEDIE 715  
DB 762 VGSSASFGARATQGLNNGVPPVSPYGOATINLP---TANKYAPVSPTV----- 808  
QY 716 ASNEEENPEDSEVAVKPSEDSPEN--ATSRGNETPAVELEP-----TTETAPSTSPSLA 768  
DB 809 ---QOKQYPSVVQNLGASAVNTPNFKVTHRGHTSSISYTPNQNEHASRYAPNTQQSQY 864  
QY 769 VPSTKP-----AESVETQVNDISISAEABQMDVDQEHSAEBSVCDPP-----PA 816  
DB 865 VPYTSQVGPVAGNSYQSSQTRSSYAVPMMPQA---QTSASIQPHANIQPPTGILPLAPL 921  
QY 817 TKADSVDEVVRVPENHASKVEGDNTERDLDRASEKVEPRDEDLVVAQOINAOPEPQSD 876  
DB 922 RPLDPLQATNL-QPRASNTAANSLPLANLPLAENILP---EITHRATSSVAPPROEN 977  
QY 877 NDSSATCSADEVDGEPERQRMFP-----MDSKPSLINPTGSIIVSS---PLKPN 923  
DB 978 NPIK-----IDNEALLRQFPFIHWSAANKVYVAPPIDQSOYMISSIVQEIKT 1029  
QY 924 PLDLPLQOHRAAVIPP--MVSCTPCNIPIGTPVSGYALYQRIKAMHESAL--LEEQROR 979  
DB 1030 PID-----QIKPNDMLKSF-----GPLGSAKLKKDLTKWMTETIKSISENESS 1075  
QY 980 QEQ-----IDLECRSSTSPCGTSKSPNREWEVL---OPAPH-QLITNL----- 1018  
DB 1076 TDMTWIQLLEKMLNDKVNWKNIKLLYNSDELLMYLSQFPFNGDMIPNAYRLDINCQMRV 1135  
QY 1019 -----PEGVRLPTRRTPRPPPLIPSSKTTTVAEKPSFIMGSGISQ--TPG----- 1063  
DB 1136 LAFLOTGNHDEALRLALSKRYAIALVGS---LMGKDRWSEVIQKLYEGFTAGPNDOK 1192  
QY 1084 -----TYLTSNQASYTOETPKPSVSGSISGLPLPROQESAKSAT 1101  
DB 1193 ELAHLFLLIFQVFGNSKMAIKSFYTNNETSQWASENKSIVAALINIPENNED--PLL 1250  
QY 1102 LPYIKQE---EFSPRSONSQPEGLLVRAHQHEGVVGTAGAQEGSITRGTPTSKTSVESI 1158  
DB 1251 IPPVVLEFLIEF-----GIFL-----TKKGLTAAASTLFIIGNVPLS----- 1287  
QY 1159 PSLRGSITQGTPALPOTGIPTAEALVKGISRMPIEDSDSPEKGREEAASKGHVYIEGKSGH 1218  
DB 1288 -----NEPVMADSDVIFE---SIGNMNTFES-----ILWDEIYEV 1319  
QY 1219 ILSVDNINKNAREGTRSPRTAHEISLKRYSVEGNIKQGM-----SMRESVPV 1265  
DB 1320 IFSYDP-KFKGFSSTLPQKIYHASLLQ-----EGLNSLGTKYTDYLSSSVYRKLPK 1369  
QY 1266 SAPLEGLICRALPGSPHSDLKERTVLSGSIQGTPTATTESFEDGLYKPKQIKRESPP 1325  
DB 1370 KDILITINTREL-----SEVASRLSESTNGWLAKPKLSSVWGQDLKFNKYGDD--I 1421  
QY 1326 RAFEGAITKGPYDITTIKEMGRSIEHPRQDILTOESRKTPEVVQSTRPIIEGSIQSG 1385



Db 1422 DALNKKNDKKVDFGFTP-----GSSANS-----STVDLTQTFPTF-----QAQV 1461  
Qy 1386 TPIKFDNNGSGLKHNKVS---LITGSKLSRGMPLLEIVPENIKVYKERYEDVRAGE 1442  
Db 1462 TSQSYVDFTALLHNAHVNSVLSHSPNSVKG-----LVEANLPYTHR--IGDSLOGS 1514  
Qy 1443 TVSRHTSVSSGSPV-----LRSTLHEAPKAQISPGIYDTSARRTPVSYONTMSRGP 1497  
Db 1515 PORIHTQFAAAEPOMASLRVRVTDQHTNEKALKSQOILEKKSTAYTPQFGON---HSPV 1571  
Qy 1498 MNRTSDV-----TIPPN--KSTNHERKSTLPTOBESIPAKSPVPGVDVP 1541  
Db 1572 MEKSNNSVPSLAFDPAPPKLGTVPNSVSPDLVRRESIISTGSEFLP---PPKIGV--- 1626  
Qy 1542 VSHSPFDPHHRGSTAGEYVWS---HLPTOLDPAMP-----F 1574  
Db 1627 -----PTKANSOGSLMSPSVEALP--IDPVVPQVHETGYNDFGNKHKSQKSWPEDES 1677  
Qy 1575 HRAIDPAAAYLFORQLSPT-----PGYPSOYQLYAME-----NTRQTI 1613  
Db 1678 HTSHDNSNAD---QNTLKDSADVTDDETMDIEGPGFNDVKLLPMEPNHQPTSTVNPQTI 1734  
Qy 1614 LNDYITSOQMOVNLR-----PDVARGLSPREOPLGL-----PYPATRGIDLT 1656  
Db 1735 SDDIPIQITNVVEVGTDAKMNLSIENRSESEOPENISKSASAYLPSTGGLSLE 1794  
Qy 1657 NMPTILVPHPGGTSTPMDRITVPGTQITFPFRPNYSASMSGPHPTHLLAAASAEER 1716  
Db 1795 NRPLT-----QDENSISQVSTVLPAGSISMEAKPIISOVDVPRNVN--KASKLVQHM 1848  
Qy 1717 EREKERERERI-----AAASDLYLRPGSEOPGRGSHGYKRSPPSV--RTQE 1765  
Db 1849 APPKPKSTDAKMNYSVPVPOSTAASAD-----GDESTILKTSPIAYARTHQ 1895  
Qy 1766 TML-----QORPSVQCGNGISVITPLDPTAQLRIMPLPAGGPSISQGLPASRYNTAADAL 1821  
Db 1896 AHASNPQYFPLVQ--ANETASFELSESTSOAQ-----SNGNVASENRFSPIKK 1943  
Qy 1822 AALVDAASAPQMDVSTKSKSAAAR--LEENLRSAVAASEQOQ-----LEOKTLEVE 1874  
Db 1944 AEVVEKDTFQ--TIRKASTNQYAFPLESDADKYNDVIEDESDDDDNMSTDEAKNKEE 2001  
Qy 1875 KRSVOCLYTSAPFSGKP-QPHSVSVTSEAGKDKGPPPKSRYEELRTRGKTTITANFI 1933  
Db 2002 KKNYN--MKKETPSNKDIIDDKSGWGLKKDTG--DKVKYAKL--GHKNTL-----YY 2051  
Qy 1934 DVITITROIASDKARERGSQSSDSSLSHRYETPDSAEIVSPASSPAPQEKLOTYQ 1993  
Db 2052 DEKLRWV--NKDATEEKQ-----KITESSAPPPPIVVRKDDG 2089  
Qy 1994 PEVYKANQAENDPTROYEGPLHRYPOQESPSPOOQLPPSPQAEGMGQVPRTHRLITLAD 2053  
Db 2090 PK-----TKPRSGPINN-----SLPPV-----2106  
Qy 2054 HICQIITQDFARNVSSQTPOOPPTS---TFQNSPVALVSTPVRTKTSNRYSPESQAQSV 2110  
Db 2107 HATSVI-----PNNITGEPLPIKTSPS-----PTGPNPNNSPSS-----2143  
Qy 2111 HHQRPGRSVSPENLVDK 2127  
Db 2144 ----PISRISGVNLTSK 2156

RESULT 38  
MAP2\_MOUSE  
ID MAP2\_MOUSE STANDARD; PRT; 1828 AA.  
AC P20357;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE MICROTUBULE-ASSOCIATED PROTEIN 2.  
GN MAP2 OR MTAP2.  
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89083571; PubMed=3205744;  
RA Wang D., Lewis S.A., Cowan N.J.;  
RT "Complete sequence of a cDNA encoding mouse MAP2.";  
RL Nucleic Acids Res. 16:11369-11370(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89043973; PubMed=3142041;  
RA Lewis S.A., Wang D., Cowan N.J.;  
RT "Microtubule-associated protein MAP2 shares a microtubule binding  
RT motif with tau protein.";  
RL Science 242:936-939(1988).  
CC -1- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY  
CC STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO  
CC SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.  
CC -1- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M21041; AAA39490.1;  
DR PIR; S06467; S06467.  
DR PIR; A40115; A40115.  
DR MGD; MGI:97175; Mtap2.  
DR InterPro; IPR001084;  
DR Pfam; PF00418; tubulin-binding; 3.  
DR PROSITE; PS00229; TAU\_MAP\_1; 2.  
DR Microtubules; Repeat; Calmodulin-binding (POTENTIAL).  
KW DOMAIN 1452 1472 CALMODULIN-BINDING (POTENTIAL).  
FT REPEAT 1662 1692 TAU/MAP MOTIF.  
FT REPEAT 1693 1723 TAU/MAP MOTIF.  
FT REPEAT 1724 1755 TAU/MAP MOTIF.  
SQ SEQUENCE 1828 AA; 198980 MW; 200BC59E360538CA CRC64;  
  
Query Match 2.3%; Score 296.5; DB 1; Length 1828;  
Best Local Similarity 19.2%; Pred. No. 0.00039;  
Matches 371; Conservative 226; Mismatches 659; Indels 667; Gaps 91;  
  
Qy 832 HASKVEGDNTKERDLDRASEKVEPRDEDLVVAQOINQAORPEPQSDNDSSATCSADEVDG 891  
Db 61 HRSQGYSDTKENGING-----ELTSADRETAEEVSARIVQ-----VVTAEAVAVLKG 108  
Qy 892 EPERQRPFPMDSKFSLNPTGSLVSPKPNPLDLQLOHRAAVIPWVSCFPCNPIG 951  
Db 109 EQEKEAQY-----KDPAALPLAAEETANLPSPPPSP-----141  
Qy 952 TPVSGYALYQRIKAMHESALLEEQROEQIDLECRSSTSPCGTSKSPNREWEVLQAP 1011  
Db 142 -----ASEQTAIVEE-----DLITAS-----KMEFFPEQEKFP 168  
Qy 1012 HQLITNLPVGLRPTTRP-----PPPLIPSSKTTVASKEPSFIMGSSIQGTP-----1062  
Db 169 SSFAEPLDKG-EMEFKMPKPGEDFEHAALVPTSKTPQDKDLQGMGEGLPPVPPFAQT 227  
Qy 1063 -GTYLTSNQASYTQETPKPSVGSISLGLPQOESAKSATLPYIKQEEFSPRSQSQEP 1121  
Db 228 FGTNLEDRKQST-----EPSIVMPSTGLSAEPAPKPKDWFIEMPTESKDE-----275  
Qy 1122 LLVRAHQEVVVRGTAGATQEGSITRGTPTSKISVETSPSLRGS-----ITQGTAL 1172  
Db 276 -----WGLAAPISPFILT--PMREKDVLEDIPRWEGKQFDSMPSPFFGGFTL 322  
Qy 1173 PQTGIPTEALYKGSISRMPIEDSSPEK-----GREAAKSHVYIEGKSGHI--LS 1221

Db 323 PLDTMKNRSVSEGPFPAPVFQSDSKYSLQDPSALATSKSSKDEEPLKADKADVDS 382  
Qy 1222 YDNKNAREGTRSPRTAEISLKRSEVEGNIGKQMSRSPVSALEGLICRALPRGS 1281  
Db 393 ISEVTTILGNVHSPVVEGVG-----ENISGEVTTTQDEKETSAP--SVQEPFLTETE 435  
Qy 1282 FHSCLKERTVLS-----GSMQGTTPRATWTFSEFDGLKYP-KQIKRE 1321  
Db 436 POTKLDKSTVSEIEBAVEESLKLDDKTGVITQTSFEQSFREDQKQOEHIDELQD 495  
Qy 1322 SPPTRAFGATKGPYDGIITIKEMGRSIIHEIPQDILLTOESRKTPEVVOSTRPIEBS 1381  
Db 496 SFPI-SLEQAVT-----DAAMTSKTLGR-----VTSE-----PEAV-SERREIOGL 534  
Qy 1382 ISOGTPIKFDNNSGOSALKHNKSLITGPKSLRGMPLIEIVPENIKVVERGKYEDYKAG 1441  
Db 535 FEETADKNKLEGGASA-----TIAEVMPFYED----- 563  
Qy 1442 ETVRSRHTSVVSSGSPVLRLSTLHEAPKAQLSPGIYDDTSARRTPVS-----YQNTM 1492  
Db 564 -----KSGMSK-----YFETSALKEDWTRSTELGSDYIELSD 595  
Qy 1493 SRGSPMMNRTSDVITPPNKSNNHERKSTLTPTQRESIPAKSPVPGVDV--SHSPFDPHH 1551  
Db 596 SRGS-AQESLDTISPKNQ--HDEKELQAKASQSPPPAQE--AGYSTLAQSYTPGHP-- 646  
Qy 1552 RGSTAGEVYWSHLPTQDAMPFHR--ALDPAAAAYLFQRLSTPTPGYPSQYQLYAMENT 1609  
Db 647 -----SELPE--BPSPQOERMFTIDPKV-----YGEKRDLSKNNK 680  
Qy 1610 RQFILNDY-----ITSQOMQVNL-----RPDVARG-----LSPREQPLGLPYPATR 1650  
Db 681 DTLRSGLGGRSAIEQORSASINLPMSCDLSIALGFNFGRHDLSPLASDI---LTWTS 737  
Qy 1651 GIIDLTN--MPTTILVHPGTSITPPNDRIYIP-----G 1683  
Db 738 GSMDEGDDYLPPT-----TPAVEKMPCFPTESKEEDKAEQAKVTGGQITQVETS 787  
Qy 1684 TQITFPPRPY--NSASMSGPHGTHLAAA-----ASAEEREREREKERERETAAASD 1735  
Db 788 SESPPFAKEYKNGTVMAPDPEMLDLGATGRSLASVADAEVARRKSVSPSEAMLAES 847  
Qy 1736 LYLRPGSQP--GRPGSH-----GY-----VRSPSPSVRTQETMLQRPVSVFGTNGTS 1782  
Db 848 SLPPVADESPTVVRPDSQLEDMGYCFVKNYTVPLSP--VQDSENLSGSGSYEGTD-DK 905  
Qy 1783 VITPLDPTAQRLMPLPAGG-----PSIS---OGLPA-----SRYNATA 1817  
Db 906 VRRDLATDLSLIEVKLAAGRAVKKDEFTEAKEATPPTSADKSGLSREFDHRKANDKLDTV 965  
Qy 1818 ADALAALVDAASAPO-----MDVSKTKE-----SKH 1844  
Db 966 LEKSEEHIDSKHAKESSEMGKVELFGITGYDOASTKELITTKDTSPEKTERGLSSVP 1025  
Qy 1845 EAARLEENLRSR-----SAAVSEQQOLE-----QKTLEVEKRSVOCLY 1882  
Db 1026 EVAEVEPTTKADQGLDPAATKAEPSQLDIKYDFQOMASGMNVNAGKAEILKFEVAQELT 1085  
Qy 1883 TSSAPPSKGPQPHSVVYSEAG--KDKGPPPKSRYEELFRGKTTTITAAFNDFVITRQ 1940  
Db 1086 LSSE-----APOEADSFMGVESGHKEGKVNETEVEKE-----VTKPLDV-----HQ 1128  
Qy 1941 IASDKARERGSQSSSSLSHRYETPSPDAIEVSPASS-----PAPP- 1985  
Db 1129 EAVDK---EESYESGHEHSLTMESLK-PDEGKETSPETSLIOEVALKLSVEIPCPPP 1184  
Qy 1986 --QEKLOT-----YQPEVVKANQANDPTROYEG-PLHHVYRPOQES-----PSPQQQ 2029  
Db 1185 VSEADLSDEKGEVQMEFIQLPKEESTETPDIPALPSDVTQPEQEAIVSEPAEVPSEEEE 1244  
Qy 2030 LPPSQEGMGQVPTRLITLADHICQITQDFARNQVSSQTQQPPTSTFQNSALY 2089

Db 1245 IEAGGEYD-----KLLFRSDTL-----QISDLLVSESEREFVETCPGELK 1284  
Qy 2090 S-----TPVTKT-----SNRYSPEQAQSVHHQRPQS-----RVSPEN 2123  
Db 1285 GVSVSVTIEDFTVTVQTTDEGESGSHSVFAAPAQAQPEERPRPHDELEIEMAAEA 1344  
Qy 2124 LVDSKRSRPG--KSPERSHVSSEPIEP-----ISPPQVYVHVKQDSLL 2166  
Db 1345 QAEPKDGSPDAPAPPEKEEVAFSEYKTTYDDYKDEITIDDSINDADSLWVDQTDDDRSI 1404  
Qy 2167 LLSORGAEPAQR--NDARSP-----GSSIS-----YLPSTFTKLENT 2201  
Db 1405 LTEQLEITPKERAEKDAARPSLEKHKRKEPKFTGCRGRISTPERKVKAKPESTVSRDEV 1464  
Qy 2202 SPVYKSKQEIFRKLNSGGGSDMAAAQPCTEIFNLPVTTSGSVSRGHSFADPASNL 2261  
Db 1465 RKKAVYKKAELAKK-----SEVQAHSPSRKILILPAI-----KYTRPTH-- 1503  
Qy 2262 GLEDIIRKALMGSEF-----DKVEDHGVVMS-----QPMGVVPGTANTSV 2301  
Db 1504 -LSCVKKRTTAASGDLAQAPGAFKQAKDKVTD-GISKSPKRSSLPPRSSILPPRG-- 1558  
Qy 2302 VTSGETRREEDPSPHSGGVCKPKLISKNSRKSIPGOGYLGCTERPSSVSY-----H 2357  
Db 1559 -VSGD--REENSFSLNS-----ISSARTRTSEPIRRAGKSGTSTPTTGGTATPG 1608  
Qy 2358 SEGDYHRTQCGWAMEDRPSSTGTFQFVYNPLTMR-----MLSSSTPTPTACAPS 2406  
Db 1609 TPYSYKRTPG-----TPGTSPYRTPGTGKSGILVPSEKKVAIIRTPKSPATPKQ 1660  
Qy 2407 --AVNOAAPHOON 2417  
Db 1661 LRLINQPLPDLKN 1673

RESULT 39  
NFM\_MOUSE  
ID NFM\_MOUSE STANDARD; PRT; 848 AA.  
AC P08553: Q61961;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NEUROFILAMENT TRIPLET M PROTEIN (160 KDA NEUROFILAMENT PROTEIN)  
DE (NEUROFILAMENT MEDIUM POLYPEPTIDE) (NF-M).  
GN NEFM OR NFM.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87246694; PubMed=3036526;  
RA Levy E., Liem R.K.H., D'Eustachio P., Cowan N.J.;  
RT "Structure and evolutionary origin of the gene encoding mouse NF-M,  
RT the middle-molecular-mass neurofilament protein.";  
RL Eur. J. Biochem. 166:71-77(1987).  
RN [2]  
RP SEQUENCE OF 322-540 FROM N.A.  
RX MEDLINE=87158637; PubMed=3103856;  
RA Julien J.-P., Meyer D., Flavell D., Hurst J., Grosveld F.;  
RT "Cloning and developmental expression of the murine neurofilament  
RT gene family.";  
RL Brain Res. 387:243-250(1986).  
CC -I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
CC -I- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFM IS  
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF  
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
CC OF AXONAL CALIBER.  
CC -I- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE  
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND



CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR  
 CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES.  
 CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM  
 CC 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N-  
 CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT  
 CC NONCOVALENTLY, ASSOCIATED.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- ALTERNATIVE PRODUCTS: A SHORT FORM MAY ARISE BY ALTERNATIVE  
 CC SPLICING IN THIS VARIANT THE N- AND THE C-TERMINAL FRAGMENTS FAIL  
 CC TO ASSOCIATE.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL TISSUES AND THE  
 CC ADULT KIDNEY. PRESENT IN ALL TISSUES TESTED.  
 CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE  
 CC SIGNAL.  
 CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH  
 CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS  
 CC AT A DEFINED SITE, PCPE/THET, WITHIN THE HCF REPEAT.  
 CC -1- PTM: GLYCOSYLATED; CONTAINS O-LINKED N-ACETYLGLUCOSAMINE RESIDUES.  
 CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: L20010; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL: X79198; CAA55790.1; ALT\_INIT.  
 CC MIM: 300019; -  
 CC InterPro: IPR001777; -  
 CC Pfam: PF00041; fn3.1.  
 CC Nuclear protein; Repeat; Alternative splicing; Glycoprotein.  
 CC  
 CC REPEAT 44 89  
 CC REPEAT 93 140  
 CC REPEAT 148 194  
 CC REPEAT 217 265  
 CC REPEAT 266 313  
 CC REPEAT 1010 1439  
 CC REPEAT 1010 1035  
 CC REPEAT 1072 1097  
 CC REPEAT 1101 1126  
 CC REPEAT 1158 1183  
 CC REPEAT 1286 1311  
 CC REPEAT 1314 1339  
 CC REPEAT 1349 1374  
 CC REPEAT 1414 1439  
 CC REPEAT 382 450  
 CC REPEAT 564 564  
 CC REPEAT 603 603  
 CC REPEAT 1164 1164  
 CC REPEAT 1873 1873  
 CC REPEAT 2035 AA; 208841 MW; 99207FBE875204C0 CRC64;  
 CC SEQUENCE

Query Match 2.3%; Score 295.5; DB 1; Length 2035;  
 Best Local Similarity 19.6%; Pred. No. 0.00049;  
 Matches 394; Conservative 226; Mismatches 780; Indels 607; Gaps 91;  
 QY 540 EKESKENTKEKIDGTAET---EREQATPRGRK-----TANQGRRK 583  
 D 287 EKEKNTTTLACLDNTMAWETILMDTLEDNIPRAGHCAVAINTRLYIWSGRDGYRKA 346  
 QY 584 RITRSMTNEAAAAAATAEPPEPPPPPEPTEPETSRTWTEEMEVAKKGLVEH 643  
 D 347 -----WNNQVCCDLWLETEKEPPPP-----ARVOLVRANTSLEV----- 382  
 QY 644 GRNWAATAKMGTKSEACKNFYKRRHNLNLLQOHKOKTSRKPREERDVSQCESVA 703  
 D 383 --SWGAVA-----TADSYLLQ-----LQYDI--PATA 406  
 QY 704 STVSAQEDIEASNEENPEDSEVAVKPSEDSPENATS--RGNTPEAVELEPTTETAP 761

Db 407 AT-----ATSPTPNPVPS--VPANPPKSPAPAAAAVAQVLTQVGTILLQAAPAP 455  
 QY 762 STSPSLAVSTKPAEDSVETQVNDSSISAETAQMDVDQEHSAEAG-----SVCDPPPA 816  
 Db 456 PTTTITQVLTVP-----GSSISVPTA-----ARTQGVPAVLKVTGPQAT 495  
 QY 817 TKADSDVDEVRVPHENHASKVEGDMTKERDLDRASEKVEPRDELDLVAQ-----INAQRP 871  
 Db 496 TGTPLVTMR-----PASQAGKAPVTVTS-----LPAGVRMVVPTQSAQGTIVIGSSP 541  
 QY 872 EPQSDNDSSATCSADEVDGEPERQRMFMDSKPSLLN--PTGSILVSS--PLKPNPLDLQP 929  
 Db 542 QMSMAALAAAAATQKI-----PPSSRPTLVSPAGTIVIKTMAVTPCTTTLPA 591  
 QY 930 LQHRAAVIPMWSTPCN-----IPGTPVSGYA--LYQRHIKAMHESALLEQRORQE 981  
 Db 592 TV-KVASSPVWVS--NPATRMKLKTAQAQVTSVSSATNTSTRPIITHKSGTIVTVAQQAQ- 648  
 QY 982 QIDLECRSSSTPCGTSKSPNREWEVLQAPHOLITNLPQGVRLPTTRP----- 1029  
 Db 649 VVTTVGGVTKTTLVXSP-----ISVPGSALISNLGKMSVVQTRPVQTSVAVTQAST 703  
 QY 1030 -----TRPPPLIPSSK--TTVASEKPSFIMGSGISQGTPTGTYLASHNQASYTQETPK 1080  
 Db 704 GPVTQIIOTKGPLPAGTILKLVTSADCKPTIITTTQASGA-GPKPTILGISSVSPSTK 762  
 QY 1081 PSVGSISGLPRQ-----QESAKSAT-LPYIKQEFSP-----RSQNSQPGELL--- 1123  
 Db 763 PGTTIITKTIPMSAITQAGATGVTSSPGK-----SPITITTKVMTSGTGAPAKIITAV 818  
 QY 1124 --VRAQH-----EGVVRGTAGIAQEGSITRGTP-----TSKISVESI--PSLRGSITQG 1168  
 Db 819 PKIATGHGQGVTVQVLKAGP--QPGTILRTVPMGGVRLVPTVTSVAVPAPVTTLVKKG 876  
 QY 1169 TPALPOTGIPTALVKGSIISRMPIEDSSPEKGRBEEAAKSHVIYEGSKHLSYDNKNA 1228  
 Db 877 T-----TGVTTLTGTVTGTVS-----TSLAGAGGHSTASLATPTITLTGIATL 919  
 QY 1229 REGTRSPRTAHEISLKRSYESVEGNIQKGMRESPPVSAPLEGICLALPRG---SPHSD 1285  
 Db 920 SSQVINP-TAITVSAQTTTLTAAGGLTT-PTITMQPVVSQPTQVTLITA-PSGVEAQVPHD 976  
 QY 1286 LKERTVLGSGIMQTPRATTESFEDGLKYPKQIKR--ESPPIRAFAEGAITKGPYDGITT 1343  
 Db 977 LPVSILASPTTEQTAIVTADSCQGDVQGTVTLVCSNPPCETHETGTNTAT---TTV 1033  
 QY 1344 IKEMGRSHEIPRQ--DILTOESRRTPEVOSTRPIIEGSIQ--GTPIKFDNNSGQSAIK 1400  
 Db 1034 VANLGG--HPQPTQVQVCDRQRAAAASLVTSTVGQONGSVVRVCSNPPCETHETGTNTA 1091  
 QY 1401 HNVKSLITGPKLSRGMPPLEI-----TPVSQNTMSRGSPMM-----NFTSD 1504  
 Db 1092 TTATSNMAGOHGCSN--PPCETHETGTNTATTAMSSVGANHOHARRACAAAGTAPVIRI 1149  
 QY 1430 -VERGKYEDYKAGE-TVRSRHTSVVSS-----GPSVLRLSTLHEAP--- 1467  
 Db 1150 SVATGALAAQSGKPCQQTQTSATSTTMTVMATGAPCSAGPLLGPSMAREPGRSAPAFV 1209  
 QY 1468 -----KAOL-SPGIYDDTISARR-----TPVSQNTMSRGSPMM-----NFTSD 1504  
 Db 1210 QLAPLSKVLRSPSIKDLPAGRHSHAVSTAATMRSSVGAGEPRMAPVCSLOGSGSSTT 1269  
 QY 1505 VTI-----PENKSTNHERKSTLTPTQRESIPAKSPVPGVDPPVVSHS 1545  
 Db 1270 VTVTAELALCPSATVTQVCSNPPCEH--HETGTNTATTATSNAGSAOR-----VCSNP 1320  
 QY 1546 PFDPHHGRST-----AGEVYWSHLPTQ-----LDPA 1571  
 Db 1321 PCETHETGTTHATTATSNNGTGOGEQGPAGRCPCETHOTTSTGTMTSVSGVALLPDA 1380  
 QY 1572 MPFHRAID-----PAAAAYLFQRLSPTP-----GYPSQVLYAMENTROTILNDY 1617

Db 1381 TSSRTVESGLEVAAPSVTPQAGTALLAPPFTQRCVSNPPCETHETGTTTATTVTSM 1440  
QY 1618 ITSQOMQNLNRPDVARGLSPREPLG-----LPYPATRGIDLTNNPPITL 1663  
Db 1441 SNQD-----PPPAASDOGEVESTQDSVNTSSAITTTVSSUTRAVTTVQSTP--- 1492  
QY 1664 VPHPGGTSTPPMDRITYIPGTQITPPPPYNSASMPGHPHAAAASAEREREREKE 1723  
Db 1493 VP--GPSVPEELQVSPGPRQLPPRQL-----LQASASTALMGESAELVSA 1537  
QY 1724 RERERIAA---SSDLYLRPSEQPRGSHGVRSPPSVRTQTML---QORSVFQ 1777  
Db 1538 SOTPELPAADVLSTGEPSSQESAGSAVAVTVVQPPPTQSEVDQLSLPQELMAEAQA 1597  
QY 1778 TNGTSVITPLDPTAQLRMPLPAGGPSISQGLPASRYNTAADAALVDAASA-----P 1832  
Db 1598 GTTILMVTGLPEE-----LAVTAAEAQAQA--AATEAQAALAIQAVLQAQAQVNGTGE 1651  
QY 1833 QNDVSKT-----KESKHEAARLEENLRSAAVSEQQOLEQKTLVEKRSVQCCLYTSSA 1886  
Db 1652 PMDTSEAAATVTAELGHLAEGQEGQATTIPVLTOQELAAALVQOQLOEAQAQOQH 1711  
QY 1887 PPSGKPPQSHSVVSEACKDKGPPKSYREBELRGTGTTTAAFNIDVITRQTASDKD 1946  
Db 1712 LPTALAPADSL-----NDPAIESNCLNELAGTVPSTVAL----- 1746  
QY 1947 AREGSSQSDSSSLSHRYETPDAIEVISPASSAPPQ-----EKLQTYQPEVVKA 1999  
Db 1747 -----LPSTATESLAPSNFTFVAPQPVVAPAKQAATLVEVA 1785  
QY 2000 NOAENDPTROYEGPLHYRPOQESPPOQ-----LPPSSQAEGMG 2040  
Db 1786 NGIESLGKVDLP-----PPSKAPMKENQWFDGVIKGTVMVTHYFLPDD----- 1834  
QY 2041 QVPRHRLITLADHICQITQDFARNQVSSQTPOQPPTSTQNSALVSPVPTKTSNR 2100  
Db 1835 AVPSDDDLGTVPDY-----NOLKKQELQ-----PGTAKYKRVAGINACA 1873  
QY 2101 YSPESQAQSVHQRGSRVSPENLVKRGSRPGKSPERSHVSSEPYPIPPQVPVYHE 2160  
Db 1874 RGFSEISAFKCLTCLGFGAPCAI-----KISKSPGAHLT---WPPPSVTSKIIIEY 1923  
QY 2161 KODSLLLLSQGAEPAEQRNDARSPGISYL-----PSFF---TKLENTSPMYKSKKQ 2210  
Db 1924 SVYLAIQSSQAGG-----LKSSTPAQLAFMRVYCGPSPCLVQSSSLSNAHIDYTTKPA 1978  
QY 2211 EIFR--KLNSGGGSDMAAQPGETE 2235  
Db 1979 IIFRIAARNEKYG-----PATQV 1997

## RESULT 41

2EPL\_HUMAN STANDARD; PRT; 2717 AA.  
AC P15822;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ZINC FINGER PROTEIN 40 (HUMAN IMMUNODEFICIENCY VIRUS TYPE I ENHANCER-BINDING PROTEIN 1) (HIV-EPI) (MAJOR HISTOCOMPATIBILITY COMPLEX BINDING PROTEIN 1) (MBP-1) (POSITIVE REGULATORY DOMAIN II BINDING FACTOR 1) (PRDI1-BF1).  
GN HIVEP1 OR ZNF40.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-90169514; PubMed-2106471;  
RA fan C.M.; Maniatis T.;  
RT "A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence.";

Genes Dev. 4:29-42(1990).  
RL [2]  
RN STRUCTURE BY NMR OF 2113-2142.  
RP MEDLINE-91064333; PubMed-2248949;  
RX Omichinski J.G.; Clore G.M.; Appella E.; Sakaguchi K.; Gronenborn A.M.;  
RA "High-resolution three-dimensional structure of a single zinc finger from a human enhancer binding protein in solution.";  
RT Biochemistry 29:9324-9334(1990).  
RL [3]  
RN STRUCTURE BY NMR OF 2087-2142.  
RP MEDLINE-92232684; PubMed-1567844;  
RX Omichinski J.G.; Clore G.M.; Robien M.; Sakaguchi K.; Appella E.; Gronenborn A.M.;  
RA "High-resolution solution structure of the double Cys2His2 zinc finger from the human enhancer binding protein MBP-1.";  
RT Biochemistry 31:3907-3917(1992).  
RL [4]  
RN FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE 5'-GGGACTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1. IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT IN T-CELL ACTIVATION.  
RL [5]  
RN SUBCELLULAR LOCATION: NUCLEAR.  
RP [6]  
RN INDUCTION: BY MITOGEN AND PHORBOL ESTER.  
RX [7]  
RN DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH ZINC-FINGER IN-BETWEEN.  
RL [8]  
RN SIMILARITY: STRONG, TO HIVEP2.  
RP [9]  
RN THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
RL [10]  
RN EMBL; X51435; CAA35798.1; -;  
DR PIR; A34203; A34203.  
DR PDB; 3ZNF; 15-JAN-92.  
DR PDB; 4ZNF; 15-JAN-92.  
DR PDB; 1BBO; 31-OCT-93.  
DR TRANSFAC; T00497; -;  
DR MIM; 194540; -;  
DR InterPro; IPR000822; -;  
DR Pfam; PF00096; zf-C2H2; 5.  
DR PRINTS; PR00048; ZINC-FINGER.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 4.  
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;  
KW Nuclear protein; Repeat; 3D-structure.  
FT DOMAIN 406 456 ZINC FINGERS.  
FT ZN\_FING 406 428 C2H2-TYPE.  
FT ZN\_FING 434 456 C2H2-TYPE.  
FT DOMAIN 803 806 POLY-SER.  
FT ZN\_FING 958 981 C2HC-TYPE (POTENTIAL).  
FT DOMAIN 2087 2139 ZINC FINGERS.  
FT ZN\_FING 2087 2109 C2H2-TYPE.  
FT ZN\_FING 2115 2139 C2H2-TYPE.  
FT STRAND 2088 2088  
FT TURN 2090 2092  
FT STRAND 2095 2095  
FT TURN 2099 2108  
FT HELIX 2109 2109  
FT TURN 2115 2116  
FT STRAND 2123 2124  
FT STRAND 2127 2135  
FT HELIX 2127 2135  
SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;

Query Match

2.3%; Score 295; DB 1; Length 2717;









|    |      |                                                               |      |
|----|------|---------------------------------------------------------------|------|
| Qy | 1222 | YDNKNARECT-RSPRTAHEISLKRSYVESVEGNIKQGMBSRESVPSPAPLEGLICRALPRG | 1280 |
| Db | 793  | KDFYKN--GTVNAPOLPEMDL-----AGTRSLASVSADAЕVARKSVY--             | 835  |
| Qy | 1281 | SPHSDLKERTVLSGSIWQGTPTRATTESFDGLKYPKQIKRESPTAFRAGAITGKGPYDG   | 1340 |
| Db | 836  | -----SETVV-----EDSRTGL-----PPVTDENHVYKTD-----                 | 861  |
| Qy | 1341 | ITTIKEMGRSI--HEIPRODILTOESRKTPPEVOSTRPIIEGSIISOGTPIKFDNNSGQS  | 1397 |
| Db | 862  | -SQLEDGYCVFNKYTV-----LPSVQDS-----ENLSGS                       | 893  |
| Qy | 1398 | AIKHNVKSLITGSKLSRGM-PPELTVPENIKVVERKYEDVKAGETVRSRHHYSVSSGP    | 1456 |
| Db | 894  | GTFYEGTD-----DKVRRDLATDLSLI--EVKLAAGARVKDEFVSDKEASAHSGDKSGSL  | 946  |
| Qy | 1457 | S-----VLRSITHEAPK--AQLSPGIYDDTSARRTPVSYONTMSRSGPM             | 1498 |
| Db | 947  | SKEFDQEKKANDRLDVLEKSEBEHADSKEHAKKTEAGDELETGLGVYEQALAK----     | 1002 |
| Qy | 1499 | MNRTSDVTIPPNNKSTNHEKRSITLTTPQRESIPAKSPVPGVDVPSHSPFDPHHRGSTAGE | 1558 |
| Db | 1003 | -----DLISITDASSKAEKGL-----SSVPEI-----                         | 1025 |
| Qy | 1559 | VYNWHLPTQLDPAMPHPRALDPAAAYLFQRQLSPTPGYPSQOYLAYMENTRQTILNDYI   | 1618 |
| Db | 1026 | -----AEVPSKKVEQGLD-----FAVQQLDVKTSDF--                        | 1053 |
| Qy | 1619 | TSQOMVNLRPDAVARGSLPREQPLGPYPATVIRGIIIDLNMPPITILVPHPG----      | 1674 |
| Db | 1054 | --GOMASGLNIDRRATE-----LKLEATQ-----DMTPSKKAQЕADAPWGVSЕGH       | 1098 |
| Qy | 1675 | MDRITYIGHQITPPPPPNYSMSGPHHTHLAAASAERERERERERERERERERERERER    | 1734 |
| Db | 1099 | MKEGTKVSETV-----KOKAKPOLVHQEAV-----DKESYESSGHESLWTMES--       | 1144 |
| Qy | 1735 | DLVLRGSEQGRPGS-----HGYVRSP-SPSVRTQETMLQORSV-----FOGTNGT       | 1781 |
| Db | 1145 | -LKADEGKKETSPESLIQDEAVKLSVEICPPAVSRAADLATERADVOMEFIQGPKEE     | 1203 |
| Qy | 1782 | S-----VITPLDPTAQLR--IMPLPAGGPSISQGLPA-SRYNT-----AADALAALVDAAA | 1829 |
| Db | 1204 | SKETPDISITPSDVAEPLHETIVSEPAEIQSEEEIEAQGEYDKLLFRSDTL-QITDGLV   | 1262 |
| Qy | 1930 | SAPQMDVSKTKSKHE-----AARLEENLRSGSAVSEQO-----LEQKTLVEE          | 1974 |
| Db | 1263 | SGAREFEVETCPSHKGVIESVVTITEDDFTIVQVTTTDEGESGSHSVRFAALEQP--EVE  | 1320 |
| Qy | 1875 | KRSVQCLYTSSAPPSGKOPHSSVYV--SEAGDKGPP-----PKSRVEEELRTRGKT      | 1925 |
| Db | 1321 | RR-----PSPHDEEFFVEEAAEAQEPKOGSPAPASPREEVALSEYK                | 1365 |
| Qy | 1926 | TITAAINFIDVIITRQIASDKARSGSSDSS-----SSLSH                      | 1964 |
| Db | 1366 | TETVDYKDETTITDSDIMDADSLWVDTDQDDRSIMTEQLETIPKEEAKЕARSSLEKH     | 1425 |
| Qy | 1965 | RYETPSDAEIVTSPASSPAPPEKLTQTOPEVVKANQENDPTROYGGPL-HYHRPQOES    | 2023 |
| Db | 1426 | RKEKPFK-----TGRGRISTPЕКRVAKPEPSTVSRDEYRRKKAYVYKAEALAKTEVQAHS  | 1480 |
| Qy | 2024 | PSQOOLPPSSQAEGMGOVPRTHLITLADHCQIITQDF--ARNOVSSOTPOQPPTISF     | 2081 |
| Db | 1481 | PSKRFILKPAKYTRPTHUSCVKRTTAAGESALAPSVFQAQKDKVSDGVTKSPKRRS      | 1540 |
| Qy | 2082 | QNSFSAALVTPVTRTKTS-----NRYSPESQAQSVHHRPGSRVSPENLVDKSRGS-----  | 2131 |
| Db | 1541 | LPRPSSIL--PPRRGVGDRDENSFSLNSSISS--SARRTTRSEPIRRACKSGTSPTTTP   | 1596 |
| Qy | 2132 | -----RPGKSPERSHVSPEYPISP--POQVYVHEKODSLLLLLSORGAEPABQ-----R   | 2179 |
| Db | 1597 | GSTAITPGTGP--SYSSRTPTGPTGPPSPRPHTPGTGPKSAILV-----PSKKVALIR    | 1648 |

```

Qy 2180 NDARSPGSIYLPSPFTTKLENTSPMVSKSKQDEIFRKLNSGGGSDMAAQP----- 2232
 :|| :
Db 1649 TPKKSPAT-----PKQLRLINQPLDLKNVSKSI-----GSTDNIKYQPKGGVQIV 1695
 :|| :
Qy 2233 TEINLPAVTTSS-GSVSSRGHSFADPASNLGLIEDIIR-----KALMGSFDDKVEDHGV 2285
 :|| :
Db 1696 TKKIDLSHVTSKCGSLKNIRH- -RPGGGRVKIESVKLDFKEKVAQKVSGLDN--AAH--- 1748
 :|| :
Qy 2286 MSQPMGVVPGTANTSVVTSGETRREEDGSPHSVGCVCKPKLISNSNRKSKSPIPGQYL 2345
 :|| :
Db 1749 -----VPGGNGVKIDSKLNFREHAKARVDHGA---EIITSPGRSS----- 1787
 :|| :
Qy 2346 GTERPSSVSSVHSEG 2360
 :|| :
Db 1788 -VASPRRLSNVSSG 1801
 :|| :

```

RESULT 44  
 MRSP\_STAAU  
 ID MRSP\_STAAU STANDARD; PRT; 1637 AA.  
 AC P80544; Q9ZF62;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE METHICILLIN-RESISTANT SURFACE PROTEIN PRECURSOR.  
 GN PLS.  
 OS Staphylococcus aureus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; Staphylococcus.  
 OX NCBI\_TaxID:1280;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE 1061;  
 RA Savolainen K., Kuusela P., Paulin L., Korhonen T.K.;  
 RT "pls, a large repeat-rich surface protein of methicillin resistant  
 Staphylococcus aureus";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP PARTIAL SEQUENCE.  
 RC STRAIN-ISOLATE 1061;  
 RX MEDLINE=96270743; PubMed=8665912;  
 RA Hilden P., Savolainen K., Tynneläe J., Vuento M., Kuusela P.;  
 RT "Purification and characterisation of a plasmin-sensitive surface  
 protein of Staphylococcus aureus";  
 RL Eur. J. Biochem. 236:904-910(1996).  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF115379; AAD09131.1; -  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; FALSE\_NEG.  
 KW Cell wall; Transmembrane; Antibiotic resistance; Glycoprotein; Repeat;  
 KW Signal.  
 FT SIGNAL 1 48 POTENTIAL.  
 FT CHAIN 49 1637 METHICILLIN-RESISTANT SURFACE PROTEIN.  
 FT DOMAIN 1301 1582 141 X 2 AA TANDEM REPEATS OF D-[SAG].  
 FT DOMAIN 1598 1603 CONSERVED IN GRAM-POSITIVE COCCI SURFACE  
 FT PROTEINS.  
 SQ SEQUENCE 1637 AA; 175473 MW; 75BE9ADB469BD309 CRC64;

```
Query Match 2.3% Score 289; DB 1; Length 1637;
Best Local Similarity 19.2%; Pred. No. 0.00066;
Matches 263; Conservative 170; Mismatches 460; Indels 476; Gaps 60;

QY 503 QQIARPSQEEKVEEKEDKAETTKKEEKK-----DEEEKDEDSEK--NTKEKDIDG 556
 :|: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |: |
```

|    |      |                                                             |                                      |                                                |      |
|----|------|-------------------------------------------------------------|--------------------------------------|------------------------------------------------|------|
| Qy | 1380 | ----                                                        | GSTSQCTP                             | KFDNNSGQSAIKHNKSLITGSPSKLSRCMPPLEIVPENIKVVERGY | 1435 |
|    |      |                                                             | :                                    | :                                              | :    |
| Db | 1075 | EGOTGEKITTTTPTINPLAGE-----                                  | KVGEGETEVTKEPVDEITFGG                | 1119                                           |      |
|    |      |                                                             | :                                    | :                                              | :    |
| Qy | 1436 | EDVKAGET-----                                               | VRSRHTSVSSGPSVLRLSTLHEAPKQLSPGI      | 1475                                           |      |
|    |      |                                                             | :                                    | :                                              | :    |
| Db | 1120 | EENVPQGKHDEFDNLPIDGTEEVPGKIGNKETGEVTV-PPVDVVTKH-GPKAG-EPEY  | 1176                                 |                                                |      |
|    |      |                                                             | :                                    | :                                              | :    |
| Qy | 1476 | YDTSARRTPVSYQNTMSRGSPMMNRTSDVTIPPKNKSNHERKSTLTQTQRESIPAKSPV | 1535                                 |                                                |      |
|    |      |                                                             | :                                    | :                                              | :    |
| Db | 1177 | TKEEIPETRVLDPMTWEGSP--DKVA-----                             | QKGENCE-KTTTTPTTNPLTGEKV G           | 1227                                           |      |
|    |      |                                                             | :                                    | :                                              | :    |
| Qy | 1536 | PG-----                                                     | VDPVWSHP-FDPHHRGSTAGEYVWSHLPTQLDPAMP | 1573                                           |      |
|    |      |                                                             | :                                    | :                                              | :    |
| Db | 1228 | EGEPTTEVTKEPIDEVNTAPTIIH--GTREET-----                       | DPNLP                                | 1264                                           |      |
|    |      |                                                             | :                                    | :                                              | :    |

```

RESULT 45
CDP_HUMAN
ID CDP_HUMAN STANDARD; PRT; 1505 AA.
AC P39880;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CCAAT DISPLACEMENT PROTEIN (CDP).
GN CUTLI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Umbilical vein;
RX MEDLINE=93250973; Pubmed=1301999;
RA Neufeld E.J., Skalniak D.G., Lievens P.M.-J., Orkin S.H.;
RT "human CCAAT displacement protein is homologous to the Drosophila
RT homeoprotein, cut.";
RL Nat. Genet. 1:50-55(1992).
CC -1- FUNCTION: PROBABLY HAS A BROAD ROLE IN MAMMALIAN DEVELOPMENT AS A
CC REPRESSOR OF DEVELOPMENTALLY REGULATED GENE EXPRESSION. MAY ACT BY
CC PREVENTING BINDING OF POSITIVELY-ACTING CCAAT FACTORS TO
CC PROMOTERS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF TRANSCRIPTS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -1- DOMAIN: ASN AT POSITION 47 OF THE HOMEBOX MAY PARTICIPATE IN
CC REGULATING DNA-BINDING ACTIVITY BY PROMOTING HOMO- AND
CC HETERODIMERIZATION.
CC -1- SIMILARITY: CONTAINS THREE CUT-REPEATS AND A HOMEBOX DOMAIN
CC SIMILAR TO THOSE OF DROSOPHILA CUT PROTEIN. CDP MAY BE THE
CC MAMMALIAN COUNTERPART OF CUT.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on
use by non-profit institutions as long as its content is in no
modified and this statement is not removed. Usage by and for commer-
entities requires a license agreement (See http://www.isb-sib.ch/announcement
or send an email to license@isb-sib.ch).

DR EMBL; M74039; -; NOT_ANNOTATED_CDS.
DR HSSP; P10037; 1A07.
DR TRANSFAC; T00100; -.
DR MIM; 116896; -.
DR InterPro; IPR001356; -.
DR Pfam; PF00046; homeobox; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR Repressor; DNA-binding; Developmental protein; Nuclear protein; Repeat;
KW Homeobox; Coiled coil; Alternative splicing.
FT Domain 7 363
FT REPEAT 7 363
FT COILED COIL (POTENTIAL).
FT CUT'-REPEAT.

```



Db 1426 AAPGEG-----PAAPTSA PPPNSSSSAPRRPSSLQSL 1459

Search completed: September 8, 2001, 14:45:03  
Job time: 330 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 14:36:58 ; Search time 28.24 Seconds  
(without alignments)  
6581.654 Million cell updates/sec

Title: US-09-522-753-11  
Perfect score: 12643  
Sequence: 1 MSSSGYPPNQGAFSTEQSR.....EREPAPLLSAQVETLSDSD 2440

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score   | Query Match | Length | ID        | Description        |
|------------|---------|-------------|--------|-----------|--------------------|
| 1          | 11536.5 | 91.2        | 2453   | 2 S60254  | nuclear receptor c |
| 2          | 5395    | 42.7        | 1047   | 2 T46489  | hypothetical prote |
| 3          | 1879    | 14.9        | 1495   | 2 S60255  | transcription co-r |
| 4          | 584.5   | 4.6         | 5327   | 2 T13564  | microtubule-associ |
| 5          | 487.5   | 3.9         | 2364   | 2 A56577  | microtubule-associ |
| 6          | 465.5   | 3.7         | 2774   | 2 A43359  | microtubule-associ |
| 7          | 447     | 3.5         | 2464   | 1 QRM3P1  | microtubule-associ |
| 8          | 444     | 3.5         | 7962   | 2 I38346  | elastic titin - hu |
| 9          | 440     | 3.5         | 5170   | 2 T15348  | hypothetical prote |
| 10         | 437     | 3.5         | 2845   | 2 T149505 | adenomatous polyo  |
| 11         | 429     | 3.4         | 2649   | 2 T51023  | hypothetical prote |
| 12         | 415     | 3.3         | 2897   | 2 B48666  | cell proliferation |
| 13         | 411.5   | 3.3         | 3924   | 2 S37431  | ankyrin 2, neurona |
| 14         | 409.5   | 3.2         | 2361   | 2 T25752  | hypothetical prote |
| 15         | 408     | 3.2         | 2187   | 2 T30826  | nascent polypeptid |
| 16         | 405     | 3.2         | 2722   | 2 T20532  | hypothetical prote |
| 17         | 405     | 3.2         | 3942   | 2 T42730  | Bassoon protein -  |
| 18         | 404     | 3.2         | 3256   | 2 A48666  | cell proliferation |
| 19         | 402.5   | 3.2         | 1110   | 2 I51116  | NF-180 - sea lampr |
| 20         | 391.5   | 3.1         | 1226   | 2 T42717  | DNA-binding protei |
| 21         | 390     | 3.1         | 1266   | 2 S15053  | hypothetical prote |
| 22         | 389     | 3.1         | 5762   | 2 A41819  | proline-rich pepti |
| 23         | 381.5   | 3.0         | 2215   | 2 T16871  | hypothetical prote |
| 24         | 380.5   | 3.0         | 2938   | 2 T42761  | Bassoon protein -  |
| 25         | 376.5   | 3.0         | 2447   | 2 T16870  | hypothetical prote |
| 26         | 376     | 3.0         | 1791   | 2 T02345  | hypothetical prote |
| 27         | 374.5   | 3.0         | 3562   | 2 A47171  | chondroitin sulfat |
| 28         | 372     | 2.9         | 2526   | 2 T20531  | hypothetical prote |
| 29         | 365     | 2.9         | 2843   | 1 RBHUAP  | adenomatous polyo  |

ALIGNMENTS

RESULT 1

S60254

nuclear receptor co-repressor N-COR - mouse  
C:Species: Mus musculus (house mouse)

C:Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 08-Oct-1999

C:Accession: S60254

R:Hoerlein, A.J.; Naeae, A.M.; Heinzel, T.; Torchia, J.; Gloss, B.; Kurokawa, R.; Ry  
Nature 377, 397-404, 1995

A:Title: Ligand-independent repression by the thyroid hormone receptor mediated by a

A:Reference number: S60254; MUID:96008539

A:Accession: S60254

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-2453 <HOE>

A:Cross-references: EMBL:U35312; NID:g1022717; PIDN:AAB17125.1; PID:g1022718

Query Match. 91.2%; Score 11536.5; DB 2; Length 2453;

Best Local Similarity 91.0%; Pred. No. 0;

Matches 2245; Conservative 79; Mismatches 101; Indels 43; Gaps 12;

|    |     |                                                                 |  |
|----|-----|-----------------------------------------------------------------|--|
| Qy | 1   | MSSSGYPPNQGAFSTEQSRYPHVSQVYTFPNTRHQQEFAVPDYRSSHLEVSQASQLLOQ 60  |  |
| Db | 1   | MSSSGYPPNQGAFSTEQSRYPHVSQVYTFPNTRHQQEFAVPDYRSSHLEVSQASQLLOQ 60  |  |
| Qy | 61  | QQQLRRRPSLLSEFHGSDRPPQERRTSYEPFHPGSPVDHDSLESKRPRLEQVSDSHFQ 120  |  |
| Db | 61  | QQQLRRRPSLLSEFHGSDRPPQERRTSYEPFHPGSPVDHDSLESKRPRLEQVSDSHFQ 120  |  |
| Qy | 121 | RVSAAVLPLVHPLPEGLRASADAKDPAGGKHEAPSPISGQPCGDDQNASPSKLSKEE 180   |  |
| Db | 121 | RISAAVLPLVHTLPEGLRSSANAKDPAGVKHEAPSPISGQPCGDDQNASPSKLSKEE 180   |  |
| Qy | 181 | LIQSMRDVREIAKVEQOILKLKKQOOLEEAAKPPPEKPPVSPPPVEQKHSRISQIYY 240   |  |
| Db | 181 | LIQSMRDVREIAKVEQOILKLKKQOOLEEAAKPPPEKPPVSPPPVEQKHSRISQIYY 240   |  |
| Qy | 241 | DENKKAEEAHKIEFGLGPKVPELYNPQSDTKVYHENIKTNQVMRKKLIILFFRRNRHAR 300 |  |
| Db | 241 | DENKKAEEAHKIEFGLGPKVPELYNPQSDTKVYHENIKTNQVMRKKLIILFFRRNRHAR 300 |  |
| Qy | 301 | KOREQKICORYDQLEAWKVKVDRIENNRKAKESKTRREYKEQFPIRQKOREQERFQ 360    |  |
| Db | 301 | KOREQKICORYDQLEAWKVKVDRIENNRKAKESKTRREYKEQFPIRQKOREQERFQ 360    |  |
| Qy | 361 | RVGORGAGLSATIAARSEHSEIIEIDGLSEQENNEKQRLSVIPPMFMDAQRVKFINM 420   |  |
| Db | 361 | RVGORGAGLSATIAARSEHSEIIEIDGLSEQENNEKQRLSVIPPMFMDAQRVKFINM 420   |  |
| Qy | 421 | NGLMEDPMKVYKDRQFMNVWTDHEKEIFKDFIQHPKNFGLIASYLERKSPVDCVLYYL 480  |  |
| Db | 421 | NGLMEDPMKVYKDRQFMNVWTDHEKEIFKDFIQHPKNFGLIASYLERKSPVDCVLYYL 480  |  |

hypothetical prote  
hypothetical prote  
ankyrin 3, long sp  
hypothetical prote  
MHC class III hist  
hypothetical Myb f  
C14B9.6 protein -  
hypothetical prote  
All-1 protein +CRE  
tpr protein - huma  
MHC class III hist  
microtubule-associ  
alpha-A-crystallin  
mucin MUC5B, trach  
ALR protein - huma  
TCOF1 protein - mo

QY 481 TKKNENKALVRRNYGKRRGRNQIARPSQEKVEEKEEDKAEKTEKKEEKKDEEKDE 540  
DB 481 TKKNENKALVRRNYGKRRGRNQIARPSQEKVEEKEEDKAEKTEKKEEKKDEEKDD 540  
QY 541 KEDSKENTKDKIDGTAETEEREQATPRGKKTANSOGRKRGKITRSMWNEAAAASAAA 600  
DB 541 KEDSKENTKDKIDGTAETEEREQATPRGKKTANSOGRKRGKITRSMWNEAAAASAAA 600  
QY 601 AAATEEPPPLPPPEPISTPEVTSRWTEEMEVAKGLVHVEGRNWAIAKMGVTKSEA 660  
DB 601 AAATEEPPPLPPPEPISTPEVTSRWTEEMEVAKGLVHVEGRNWAIAKMGVTKSEA 660  
QY 661 QCKNFYFNKRRHNLNLLQHKOKTSKPREEDVSOCEVASTVSAOEDIEASNEE 720  
DB 661 QCKNFYFNKRRHNLNLLQHKOKTSKPREEDVSOCEVASTVSAOEDIEASNEE 720  
QY 721 ENPEDSE-----VEAVKPSDESPENATSRGNTBPVALEPTTETAPSTS 764  
DB 721 ENPEDSE-----VEAVKPSDESPENATSRGNTBPVALEPTTETAPSTS 764  
QY 765 PSLAVPSTKPAEDSVETQVNDSTSAETAQMDVDQOEHSAEESVCDPPPATKADSDV 824  
DB 765 PSLAVPSTKPAEDSVETQVNDSTSAETAQMDVDQOEHSAEESVCDPPPATKADSDV 824  
QY 825 EVRYPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQOINQORPEQSDNDSSATCS 884  
DB 825 EVRYPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQOINQORPEQSDNDSSATCS 884  
QY 885 ADEDVGEPEQRQMPMDKPSLLNLTGSIIVSSPLKPNPLDLQLQHRAAIVPPMVST 944  
DB 885 ADEDVGEPEQRQMPMDKPSLLNLTGSIIVSSPLKPNPLDLQLQHRAAIVPPMVST 944  
QY 945 PCNPIGTPVSGYALYQKHAKMESALLEQORQORQOVIDLECRSSSPCSTSKSPNREW 1004  
DB 945 PCNPIGTPVSGYALYQKHAKMESALLEQORQORQOVIDLECRSSSPCSTSKSPNREW 1004  
QY 1005 EVLPAPAHQITNLPPEGVRLPTTRTPPPPLIPSSKTTVAASEKPSFTMGSGISQGTPT 1064  
DB 1005 EVLPAPAHQITNLPPEGVRLPTTRTPPPPLIPSSKTTVAASEKPSFTMGSGISQGTPT 1064  
QY 1065 YLTSHNOASYTOETPKPSVSGISLGLPQOESAKSATLPYIKOEEFSPRSONSQPEGLLV 1124  
DB 1065 YLTSHNOASYTOETPKPSVSGISLGLPQOESAKSATLPYIKOEEFSPRSONSQPEGLLV 1124  
QY 1125 RAQHEGVVGTAGALOGESITRGTPTKISVESITPSLRGSIQTCTPALPQIGITEALVK 1184  
DB 1125 RAQHEGVVGTAGALOGESITRGTPTKISVESITPSLRGSIQTCTPALPQIGITEALVK 1184  
QY 1185 GSISRMPIEDSSPEKGREEAASKGHVIEGKSGHILSYDNITKNAREGTRSPRTAHEISLK 1244  
DB 1185 GSISRMPIEDSSPEKGREEAASKGHVIEGKSGHILSYDNITKNAREGTRSPRTAHEISLK 1244  
QY 1245 RSYESVENIKQGMRESVPSAPLEGICRALPRGSPHSDLKERTVLSGIMQGTPTAT 1304  
DB 1245 RSYESVENIKQGMRESVPSAPLEGICRALPRGSPHSDLKERTVLSGIMQGTPTAT 1304  
QY 1305 TESPEDEGLKYPKQIKRESPPITRAFEAGITKGPYDGTITTKEMGRSITHEIPRODILTOES 1364  
DB 1305 TESPEDEGLKYPKQIKRESPPITRAFEAGITKGPYDGTITTKEMGRSITHEIPRODILTOES 1364  
QY 1365 RKTPEVQSTRPIIEGSIQGTPIKFDNNSGQSAIKHNKVSILITGPKSLRGMPPLEIVP 1424  
DB 1365 RKTPEVQSTRPIIEGSIQGTPIKFDNNSGQSAIKHNKVSILITGPKSLRGMPPLEIVP 1424  
QY 1425 ENIKVVERGKYEDYKAGETVRSRHTSVVSGSPVLSRTLHEAPKAQLSPGIYDDTSARRT 1484  
DB 1425 ENIKVVERGKYEDYKAGETVRSRHTSVVSGSPVLSRTLHEAPKAQLSPGIYDDTSARRT 1484  
QY 1485 PVSQNTMSRSGSPMMNRTSDVTIPPNKSTNHERKSTLTPTORESIPAKSPVPGVDVIVSH 1544  
DB 1485 PVSQNTMSRSGSPMMNRTSDVTIPPNKSTNHERKSTLTPTORESIPAKSPVPGVDVIVSH 1544  
QY 1494 PVSQNTMSRSGSPMMNRTSDVTS--SSKASHERKSTLTPTORESIPAKSPVPGVDVIVSH 1551  
DB 1494 PVSQNTMSRSGSPMMNRTSDVTS--SSKASHERKSTLTPTORESIPAKSPVPGVDVIVSH 1551

QY 1545 SPFPDHRHGSTAGEVYWSHLPTQLDPAMPFRHALDPAALYLFQRLSPTPGYSQYQLY 1604  
DB 1545 SPFPDHRHGSTAGEVYWSHLPTQLDPAMPFRHALDPAALYLFQRLSPTPGYSQYQLY 1604  
QY 1605 AMENTROTILNDYITSOQMVNLRPDVARGUSPREQPLGLPYPATRGLIIDLITNMPPTILV 1664  
DB 1605 AMENTROTILNDYITSOQMVNLRPDVARGUSPREQPLGLPYPATRGLIIDLITNMPPTILV 1664  
QY 1665 PHPGGTSTPPMDRITYIPGTOITPPRPYNASMSPGHPTLHAAAASAEEREREREREK-- 1722  
DB 1665 PHPGGTSTPPMDRITYIPGTOITPPRPYNASMSPGHPTLHAAAASAEEREREREREK-- 1722  
QY 1723 -----ERERERIAAASDLYLRPGSEQGRPGSGHGYVSPSVTQTMLOQORSV 1774  
DB 1723 -----ERERERERERERIAAADLYLRPGSEQGRPGSGHGYVSPSVTQTMLOQORSV 1774  
QY 1775 FQGTNGSVITPLDPTAQLRIMPLPAGGSPISQGLPASRYNTAADAALVDAASAAPQM 1834  
DB 1775 FQGTNGSVITPLDPTAQLRIMPLPAGGSPISQGLPASRYNTAADAALVDAASAAPQM 1834  
QY 1835 DYSKTKESKHEAARLEENLRSRAVSEFQOOLEQKTLVEKRSVQCLYTSFAFSGKPPQ 1894  
DB 1835 DYSKTKESKHEAARLEENLRSRAVSEFQOOLEQKTLVEKRSVQCLYTSFAFSGKPPQ 1894  
QY 1895 HSSVYVSEAGKDKGPPPKRSRYEEELTRGKTTITAANFIDVITRQIASDKDARERGQS 1954  
DB 1895 HSSVYVSEAGKDKGPPPKRSRYEEELTRGKTTITAANFIDVITRQIASDKDARERGQS 1954  
QY 1955 SDSSSLSHRYETPSDAIEVISPASSPAPPEKQIYQPEVVKANOENDPTRQYEGPL 2014  
DB 1955 SDSSSLSHRYETPSDAIEVISPASSPAPPEKQIYQPEVVKANOENDPTRQYEGPL 2014  
QY 2015 HHYPQOESPSPOQ--LPPSSQAEGMGQVPRHRLITLADHICQIITQDFARNQVSSQT 2072  
DB 2015 HHYPQOESPSPOQ--LPPSSQAEGMGQVPRHRLITLADHICQIITQDFARNQVSSQT 2072  
QY 2073 PQQPTSTFQNSPALSVPVTRKTSNRYSPESAQSVHHQRPGRSVSPENLVKDSRGS 2132  
DB 2073 PQQPTSTFQNSPALSVPVTRKTSNRYSPESAQSVHHQRPGRSVSPENLVKDSRGS 2132  
QY 2133 PGKSPERSHVSSEYEPISPPQVPVHVEKQDSLLSLLSORGAEPAEORNDASPSGISYLP 2192  
DB 2133 PGKSPERSHVSSEYEPISPPQVPVHVEKQDSLLSLLSORGAEPAEORNDASPSGISYLP 2192  
QY 2193 SFTFKLENTSPMVKSKKQEIFRKLNSGGSDMAAQAQGTIEFNLPVTTSGSVSSRGH 2252  
DB 2193 SFTFKLENTSPMVKSKKQEIFRKLNSGGSDMAAQAQGTIEFNLPVTTSGSVSSRGH 2252  
QY 2207 SFTFKLENTSPMVKSKKQEIFRKLNSGGSDMAAQAQGTIEFNLPVTTSGSVSSRGH 2266  
DB 2207 SFTFKLENTSPMVKSKKQEIFRKLNSGGSDMAAQAQGTIEFNLPVTTSGSVSSRGH 2266  
QY 2253 SFADPASNLGLEDIIRKALMGSDFKVEDHGVMSQPMGVVPGTANTSVTSGTTRREG 2312  
DB 2253 SFADPASNLGLEDIIRKALMGSDFKVEDHGVMSQPMGVVPGTANTSVTSGTTRREG 2312  
QY 2313 DPSPHSGVCKPKLISKNSRKSIPGQYLGTERPSSVSSVHSEGDYHROTPGNAWE 2372  
DB 2313 DPSPHSGVCKPKLISKNSRKSIPGQYLGTERPSSVSSVHSEGDYHROTPGNAWE 2372  
QY 2372 DRPSTGTOFFPYNPLTWRMLSSPTTPIACAPSAVNOAAPHQONRIWEREPAPLLSAQY 2432  
DB 2372 DRPSTGTOFFPYNPLTWRMLSSPTTPIACAPSAVNOAAPHQONRIWEREPAPLLSAQY 2432  
QY 2433 ETLSDDSD 2440  
DB 2433 ETLSDDSD 2440  
QY 2446 ETLSDDSD 2453  
DB 2446 ETLSDDSD 2453

RESULT 2

T46489  
hypothetical protein DKFp434M075.1 - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000

C;Accession: T46489

R;Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23035

A:Accession: T46489

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1047 <AAA>

A:Cross-references: EMBL:AL137641

A:Experimental source: adult testis; clone DKF2p434M075

C:Genetics:

A:Note: DKF2p434M075.1

Query Match 42.7%; Score 5395; DB 2; Length 1047;  
Best Local Similarity 99.6%; Pred. No. 3.3e-202;  
Matches 1043; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 1394 SGQSAIKHNKSLITGSKSRGMPPLIEIVPENIKVVERGYEDVKAGETVRSRHTSVVS 1453
DB 1 SGQSAIKHNKSLITGSKSRGMPPLIEIVPENIKVVERGYEDVKAGETVRSRHTSVVS 60
QY 1454 SGPSVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRSPMMNRTSDVTIPPKNST 1513
DB 61 SGPSVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRSPMMNRTSDVTISSNKT 120
QY 1514 NHERKSTLTPQRESIPAKSPVGVDPVSHSPDPHRRGTAAGEVYWSHLPTQLDPAMP 1573
DB 121 NHERKSTLTPQRESIPAKSPVGVDPVSHSPDPHRRGTAAGEVYWSHLPTQLDPAMP 180
QY 1574 FHRALDPAALAYLFQRLSPTPPGYPSOYLXAMENTQTLINDYITSOOMQVNLRPDVAR 1633
DB 181 FHRALDPAALAYLFQRLSPTPPGYPSOYLXAMENTQTLINDYITSOOMQVNLRPDVAR 240
QY 1634 GLSPREQPLGLPYATRGIDLTNNPPTILVPHPGSTTPPMDRITYTPGQITFPFRPY 1693
DB 241 GLSPREQPLGLPYATRGIDLTNNPPTILVPHPGSTTPPMDRITYTPGQITFPFRPY 300
QY 1694 NSASMSGCHPTHAAAAASAERERERERERERERERERERERERERERERERERERER 1753
DB 301 NSASMSGCHPTHAAAAASAERERERERERERERERERERERERERERERERERERER 360
QY 1754 VRSPSVRTQETMLQORPSVFQGTNGTSVTITPLDPTAQLRIMPLPAGGSPISQGLPASR 1813
DB 361 VRSPSVRTQETMLQORPSVFQGTNGTSVTITPLDPTAQLRIMPLPAGGSPISQGLPASR 420
QY 1814 YNTAADALALVDAASAPQMDVSKTESKESHEAARLEENLRSAAYSEQQOLEQKLTLEV 1873
DB 421 YNTAADALALVDAASAPQMDVSKTESKESHEAARLEENLRSAAYSEQQOLEQKLTLEV 480
QY 1874 EKRSVQCLYTSSAPPSPGKPPHSSVYSEAGKDKGPPPKSRYEELRTRGKTTITAAFI 1933
DB 481 EKRSVQCLYTSSAPPSPGKPPHSSVYSEAGKDKGPPPKSRYEELRTRGKTTITAAFI 540
QY 1934 DVIIITROIASDKDARERGSQSSSSLSHRYETPDSATIEVISPASSPAPQEKLOTYQ 1993
DB 541 DVIIITROIASDKDARERGSQSSSSLSHRYETPDSATIEVISPASSPAPQEKLOTYQ 600
QY 1994 PEVYKNAQENDPTROYEGPLHHYRPOQESPPOQLPPSSQAEGMGQVPRTHRLITLAD 2053
DB 601 PEVYKNAQENDPTROYEGPLHHYRPOQESPPOQLPPSSQAEGMGQVPRTHRLITLAD 660
QY 2054 HICQIITODFARNQVSSQTPOQPTSTFQNSPSALVTPVTKTSNRYSPESQAQSVHHQ 2113
DB 661 HICQIITODFARNQVSSQTPOQPTSTFQNSPSALVTPVTKTSNRYSPESQAQSVHHQ 720
QY 2114 RPSGRVSPENLVDSKRSRPGKSPERSHVSSEPEYIPSPQVPPVHKEQDLSLLLSORGA 2173
DB 721 RPSGRVSPENLVDSKRSRPGKSPERSHVSSEPEYIPSPQVPPVHKEQDLSLLLSORGA 780
QY 2174 EPABQRNDARSPGISYLPSPFFTKLENTSPMVKSKKEIFRKLNSGGGSDMAAOPGT 2233
DB 781 EPABQRNDARSPGISYLPSPFFTKLENTSPMVKSKKEIFRKLNSGGGSDMAAOPGT 840
QY 2234 EIFNLPAVTTSGSVSRGHGFADPASNLGLEDITIRKALMGSFDDKVEDHGVVMSQPMGVV 2293
DB 2293 EIFNLPAVTTSGSVSRGHGFADPASNLGLEDITIRKALMGSFDDKVEDHGVVMSQPMGVV 2293
```

```
DB 841 EIFNLPAVTTSGSVSRGHGFADPASNLGLEDITIRKALMGSFDDKVEDHGVVMSQPMGVV 900
QY 2294 PGTANTSVVTSGETRREREGDPSPHSGGVCKPKLISKSNRSKSPIPGQGYLGTERRPSV 2353
DB 901 PGTANTSVVTSGETRREREGDPSPHSGGVCKPKLISKSNRSKSPIPGQGYLGTERRPSV 960
QY 2354 SSVHSEGDYHRQTPCWAWEDRPSSTGSTQFPYNPMLTMRMLSTPTPIACAPSAYNQAP 2413
DB 961 SSVHSEGDYHRQTPCWAWEDRPSSTGSTQFPYNPMLTMRMLSTPTPIACAPSAYNQAP 1020
QY 2414 HQQNRWEREPAPILLSAQYETLSDD 2440
DB 1021 HQQNRWEREPAPILLSAQYETLSDD 1047
```

# RESULT 3

S60255

transcription co-repressor SMRT - human

C:Species: Homo sapiens (man)

C:Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 05-Nov-1999

C:Accession: S60255

R:Chen, J.D.; Evans, R.M.

Nature 377, 454-457, 1995

A:Title: A transcriptional co-repressor that interacts with nuclear hormone receptors

A:Reference number: S60255; MUID:96008552

A:Accession: S60255

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1495 <CHE>

A:Cross-references: EMBL:U37146; NID:g1045654; PIDN:AAC50236.1; PID:g1045655

Query Match 14.9%; Score 1879; DB 2; Length 1495;  
Best Local Similarity 35.9%; Pred. No. 1.4e-65;  
Matches 572; Conservative 217; Mismatches 56; Indels 268; Gaps 76;

```
QY 993 PCGTS--KSPNREVEVLQAPAH----QLITNLPCEGVRLP-----TTRTRPPPPPLIPSSK 1041
DB 25 PCWTSGLPFPVPPREVAKSPHAPDPSAFYAPGCHPLPLGLHDTARVLPRPPTISNPP 84
QY 1042 TTVASEK--PSFI--MGSISOGTPTGYTSHNQASYTQETPKPSVGSISLGLPQOESAK 1098
DB 85 PLISSAKHPSVLERQIGAISOG-----MSVOLHVYPYSEHAKAP-VGPVTMGLPLPMDPKK 138
QY 1099 SATLPYIKOEFSRSONSQPEGLLV-RAQHEGVVGRGA-GAIOGSLTRGTPTTSKISVE 1156
DB 139 LAPFSGVQKQLSPRGAGPPESLGVPTAQEASVLRGTALGSPVGSITKGIPTRVPSD 198
QY 1157 SIPLRGSIQTCTPALPOTGIPTEALVKGSISRMPIEDSSP--EKGREAAASKGHVIEG 1214
DB 199 SAITYRGSIHTGTA-----DVLYKGTITRIIGEDSPSRLDGRGDSLPKGVHIEG 250
QY 1215 KSHILSYDNTKNA-----REGTRSPPTAHEISL-KRSYSEVGNIKOGMSRSPVAP 1268
DB 251 KGHVLSVEGGMVSTQCSKEDGRSSGPPHETAAAPKRTYDMMEGRVGRAIS-----SAS 304
QY 1269 LEGLICRALP--RCSPSDLKERIVLSIMOGTTPRATTESEFDGL-KYPQIKRES--- 1322
DB 305 IEGLMGRAIPPERHSPH-HLKEQHHRIGSITQGIIPRSVYEAQEDYLREAKLLRGTPP 363
QY 1323 --PIPAFEGAI-----TKGKP-YDG-ITTIKEMGRSIIHEIPRODILTQESRKTPEW 1371
DB 364 PPPSRDLTEAYKQALGPLKPKAHEGLVATVKEAGHSIIHEIPREL-----RHTELP 418
QY 1372 QSTRPIIEGSIQGTPIKFDNN-SQSAIKHNKSLITGSKSRGMPPLIEIVPENIKV 1430
DB 419 LAPPLKEGSIQTCTPLKYDTGASTTGSKKHVRSLIGSPGRTFPVHPDVMAD-ARAL 477
QY 1431 ERGYEDVKAGETVRSRHTSVVSGSPVLRST---LHEAPKAQLSPGIYDDTSARPTVS 1487
DB 478 ERACYE-----ESLKRPGTASSGGSIARGAPVIVPELGRPROSPLTYEDHGA-----P 527
QY 1488 YQNTMSRSGSPMMNRTSDVTIP----PNKSTNHERKSTLTPTQRESIPAKSPVGVDPVVS 1543
```







RESULT 5  
A56577  
microtubule-associated protein MAP 1B - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 16-Feb-1997  
C:Accession: A56577  
R:Zauner, W.; Kratz, J.; Staunton, J.; Feick, P.; Wiche, G.  
Eur. J. Cell Biol. 57, 66-74, 1992  
A:Title: Identification of two distinct microtubule binding domains on recombinant rat  
A:Reference number: A56577; MUID:92347374  
A:Accession: A56577  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2364 <ZAU>  
A:Cross-references: GB:X60550  
A:Experimental source: brain  
A:Note: nucleotide sequence not given; conceptual translation not complete  
C:Superfamily: microtubule-associated protein MAP1B

Query Match 3.98; Score 487.5; DB 2; Length 2364;  
Best Local Similarity 19.18; Pred. No. 2.3e-11;  
Matches 494; Conservative 360; Mismatches 946; Indels 793; Gaps 117;

Qy 71 LLSFHPGSDRPOERRTSYEPFHPGSPVDHDSLESKRPRLEQVSDSHFORV---SAAVL 127  
Db 79 LLSITHPAN-----KASLTLCF-----EGDWKNSLDNRHLQDFINIKLNSASIL 125  
Qy 128 PLVHPLPGLRASADAKDPAGGKHEAPSPISG-----OPC-----GDD----- 168  
Db 126 PEMEGLSEFTYLSSEVSPS---PFDILEPTSGGFLKSKPCYIPPGGRGSALFAV 182  
Qy 169 -----QNASPSKLSKEELQSDRVD-----RETAKEVQOI-- 199  
Db 183 NGFNMLINGGSEKSCFWKLIRHLDRVDSILLTHIGDNLPGINSMQKRLAELESQ 242  
Qy 200 -----LKLKKQO-----OLEEEAKRPP- 219  
Db 243 STNSNDMMKNLISPLDGVVFLNVPENLKNPEKNIMKSTEEACFTLQYLNKLSMKREPL 302  
Qy 220 -----EKPV-----SPPVQOKHRISVQIIVDNRKKAEEAHKIFEG 256  
Db 303 FRVSGNATEPVILFQKMGVGLKMYLVNPKVSSKEMQYFMQQTGNTKDKAE---LILP 358  
Qy 257 LGPKVELPL--YNPDSDFKVYHENIKTNQVMR-----KKLILFKRR 296  
Db 359 NGQVEDIPISYLA SVSSLIIVHPNAPAEKILRVLPFGNSTOYNILEGLEKUKHLDFLQ 418  
Qy 297 NHARKO-----RBQICORYDQLMFAWEKVKVDRIENPRRKAKESKT-REYIEKQ 345  
Db 419 LATQKDLTGQVSTPPVKQVCLKRAD-----SRESLKPATKPLSSKSVRKESKEE 468  
Qy 346 FPETRKQEQERQVQGRAGLSATTARSEHSEIIDLGSQENNEKOMRQLSVIPP 405  
Db 469 APEATK-ASQVEKTPKVESKE---KVIVKDKPKGVESKPSVTEKEVPSKEQS-----P 519  
Qy 406 MMEFAEQRRVKFINMGLMEDPMKVIKQDFQPMVNTDEKEIFKDKFTQHPKNGLIASY 465  
Db 520 VKAEVAERAA-----TESKPVTKDKVVKK----- 544  
Qy 466 LERSVPCVLYVYLTKKENYKALVRNRYGKRRGRNQOIAARPQOEKVEE--KEEDKAE 523  
Db 545 -EIKTKE-----EKKEPKKEVAKKEDKTLPKKDE--KPKKEAKKEIKKEE 594  
Qy 524 KTEKKEEKDEEKEDESKENTKEDKIDGTAETEEERQATPRGRKTYANSQGRKG 583  
Db 595 KKLKKEVKEKTEPLDKAKKEVKKDEKVEK---KEEKEPKKEIK----- 635  
Qy 584 RITRSMTEAAAAAATAATEEPPPLPPPPPEPTEPSTETPSWTET-EMEVAKKGLIVE 642  
Db 636 KJSDIKKSTPLSTKPKPAALK---PKVAKKEETPKKEPIAAGLKDKGRKVIKK----- 688  
Qy 643 HGRNWAATAKMVGTKSEAQCKNFYFNKRRHNLNLLQHQKQKTSRKPREERDVSQCESV 702

Db 689 EGTKTEAATAVGTAAVAAAAGV-----AASGPAKE-----LEAE 723  
Qy 703 ASTVSAQDEDEIASNEENPEDSEV-EAVKPSDESDPENATSRNTEP-----AVELEPTT 757  
Db 724 RSLMSS--PEDLTQDFBELKAAEEDVARDIKPOLLEIEDEEKLKETEFGAEVYIQKETE 781  
Qy 758 ETAPSTSPSLAVPSTK-----PADESVEVQVNSI----- 788  
Db 782 SKGAESPDEGITTEGECEQTEPELEPVKEGVGDVDDIEKFDEGAGFESSAGDYEE 841  
Qy 789 --SAETAQMDVDQOEHSAEESVCDPPPATKADSDVDVEVVPENHASKVGDNTKERDL 846  
Db 842 KAETEAEEPEDEGEDNVSGSASKHSPTEDEEIAKAAADVHIKRESVASDDDAEDM 901  
Qy 847 DRASEKVEPRDELVAQQINAAPEQSDNDSSATCSADEVDGEPERQRMFMDSKPS 906  
Db 902 DEALEKGE-----AEQSEEEGEEDKAEDAREE-DHEPDKTE--AEDYVMA 945  
Qy 907 LLNPTGSLVSS---PLKPNPLDLPOLQHRAAVIPPMVSCCTPCNIPITGTPVSGYALYQRH 963  
Db 946 VVDKAAEAGVTEDQYDFLGTAKPGVQS-----PSEPAS----- 981  
Qy 964 IKAMHESALLERQORQOBIDLECRSTSPGTSKSPNREWEVLQAPAHQILNLPEGVR 1023  
Db 982 --SIHDETL-----PGGSESEATASDEENRE---DQPEEFTATSGVTQSTI 1022  
Qy 1024 LPTRPT-----RPPPLIPSSKTTVASEKPS--FIMGGSISQGTPTGYTLTSHNQASVQ 1076  
Db 1023 EISEPTPMDEMTPRDVMYTDETNNEETESPQEFV-----NITKYESSLSQ 1070  
Qy 1077 ETPKPSVGSISGLP-----RQOESAKSATLP--YIKQEEFSPRSQNSQPEGLLV 1124  
Db 1071 EYSPVVASFN-GLSDGSKTDATDGRDYNASASISPPSMEEDKFKSALRD-----AY 1124  
Qy 1125 RAQHEGVVGRGTAGIQEISITRGTPTSKISVESIP-----SLRGSITQGTAL 1172  
Db 1125 RPEETDVKTAGELDIKDVSDERLSPAKSPSLSPSPSPPIEKTPLGERSVNFSLTPNEIKA 1184  
Qy 1173 POTGIPTAALVKGISIRMPED--SSPEKGREAAASKGHVIEKSGHILSYD---NKN 1227  
Db 1185 SAEGEAT-AVYSPGVTQAVVEEHCAEPEEKTLEVVSPSQSV-TGSAGHTPYQSPDDEKS 1242  
Qy 1228 AREGTRSPRTAHEISLKRSY-ESVEGNIKQGMRESVSPAPLEGLICRALPRGSPHSDL 1286  
Db 1243 SHLTETENAAQAVPVSEFTEAKDENERSISPMDEV-----PDSES 1286  
Qy 1287 KERTVLSGSIQGTPTARTES-FEDGLKYPKOI---KRESPIRAFEGAITKGPYDGIT 1342  
Db 1287 PIEKVLSP--LRSPPLIGSESAYEDFLSADDKALGRRESPE---FEGKNGKQGFSDRES 1340  
Qy 1343 TIKEMGRSIEH-----IP-RODILLTOESRKTPEV--QSTRPIEGSI-SQGTPI 1388  
Db 1341 PVSDLTSLDYODKOEKRAKGFIPKEDEFSPEKKAADAIMSQSALALDERKLGDSGPT 1400  
Qy 1389 KFDNNSGOSALKHNKSLITGTSKLSRGMPLLEIVPENIKVVERKYEDVAGETVRSRH 1448  
Db 1401 QVD-VSQFGSKEDTKMSISEGTVSDKSATP-----VDEGAEDTYSHMEGVASVS 1449  
Qy 1449 TSVVSSGSPVLRLSTLHEAPKAQLSPGIY-----DDT---SARRTPVSVONTMSR 1494  
Db 1450 TA-----SVATSSFPPEPTDDVSPSLHAEVSGPSHSTEDDLSVSVVQTPTTQET--- 1500  
Qy 1495 GSPMNRSTDTIIPNKNSTNHERKSTLTPTQRESIPAKSPVGVDPVVSHPFDPHGRS 1554  
Db 1501 -----EMSPSKE-ECPRPMISIPDFSPKTSKRTSPVQD-----HRSE 1537  
Qy 1555 TAGEVYVSHLPTQLDPAHPFHR-ALDPAAAAAYLFQRLSPPTGYPSS--QYOLYAMENRQ 1611  
Db 1538 Q-----SSMSIEFGQESPEHSLAND-----FSRQ---SPDHPTVCAGMLHITENGT 1581  
Qy 1612 TILNDYITISQOMQVNLRPDVARGLSPREQPLGLPYPATRGYIIDLTPMPTTILVPHGGTS 1671



Db 731 SYIQETIPGYSETEQIISDEEIHDEPDERPAPRFFSTYDLSGPEG-----PGPFASQ 786  
QY 818 KADSV-----DVEVRV-----ENHASKVEGDNTKRDLDRASEKVEP 855  
Db 787 AADSAVASSKTYGAFETELTYPPNVAAPLAEEHVSSA-----TSITECDKLSSFATS 842  
QY 856 RDELDVVAQIQINARPE-----POSNDNS-----SATCSAED 888  
Db 843 VAEDQSVASLTAPOTEETGKSSLLDVTSPSSRTEATQGLDYVPVSGAGTSPSSLEED 902  
QY 889 -----VDGEPE-----RORMFPMDSKPSLLNPTGSLIVSSP 919  
Db 903 KGFKSPCCDFSVTGESEKKGETVGRGLSGEKAIVGEKVVVTSEK--LSCQYAAVFGAP 960  
QY 920 ---LKNPPLDLPOLQHRA-----AVIPPMVSCYPCNIPIGTIPVSGYALYQRHIKAM 967  
Db 961 GHTLPPCEPALGEVEERCLSPDDSTVMASPP-----PSGPPSAAHYTF----- 1004  
QY 968 HESALLEEQRQRBQIDLECKRSSTSPCGTSKSPNREWEVLQAPAHQILITNLPEGVRLPTT 1027  
Db 1005 HQSPVEDKSEPRDFQEDSWGETHSP--GVSKEDSEE--QTVKPGPEEGTSE--EGKGPPT 1060  
QY 1028 RPTRRPPP-----LIPSSKTVASEKPSFIMGSSISQGT--PGTYLTSNHOAS 1073  
Db 1061 SPQAQDMPVSIAGQGTCTIQLLPEQDKAIVFETGE--AGSNLGAAGTLPGEVRTSTEAT 1118  
QY 1074 YTOETPKPVSIGLPLPROQESAK--SATLPYIKQEEFSPRS-----ONSQPEGLL---- 1123  
Db 1119 EPQKDEVLRFTDQSL--PEDAESLVSIVSPDTTKQENTPRSPCSLKEQOPHKDLMPV 1177  
QY 1124 -----VRAQHEGVVRGTA-----GAIQEGSITRG-----PTSK----- 1152  
Db 1178 SPEDQSLSFSEERSPKETSILDISKQSPESLGTQLQFELNLGKEERGPMKAEDDSCH 1237  
QY 1153 ISVESIPS--LRGSTQGTTPALPQTGITEALVKGSIS----- 1188  
Db 1238 LAPVSIPEPHRATVSPSTDETPACTLPGGFSHLSALSDVRKHSFGEITPGGHPMTSDSS 1297  
QY 1189 -----RMPIEDSPKEGREAASKGHVYEGKSHLSYDNIK 1236  
Db 1298 LTKSPESLSPAMEDLAVEWEGKAPGKEPELAKSETRQKQGLPKVA--VVEQDLII 1355  
QY 1227 NAREGT-----RSPRTAHEISLKRYSVESGNKQKMSRESVPSAPLEGILICAL 1277  
Db 1356 HOKDGALEENKPGQODKTPQKGRDLDEKDTAAELDKGPEKPKDLREDQ----- 1409  
QY 1278 PRGSPHDLKERTVLSIMQGT-----PRATTESFEDGLKYPKQIK--RESPPTRAFEGAI 1332  
Db 1410 QRAGPPAEKDKASQRTDLQQTQATEPRDRAQERDSEKDKSLELRDRTP----- 1461  
QY 1333 TKGKPYDGIITIKEMGRSIIHEIPRODILTQESR-----KTPPEVQSTRIPIEGSISQGTPI 1388  
Db 1462 -----EKDR-----ILVQEDRAPEHSIPEPTQDTR-----APDKGTDD 1496  
QY 1389 KFDNNSQSAIKHNV-----KSLITGPSKLSRGMPPLIIVPENIKYVERKGYEDVKAGETVR 1445  
Db 1497 K-EQKEASEKEQVLEQKDWALKEG-----ETLDQEARTEAQ--KDETLKEDKT-Q 1545  
QY 1446 SRHTSVVSGSPVLRLTLHAPKAQLSPGIYDDTSA---RRTPVSYONTMSRGSPPMNR 1502  
Db 1546 GQKSFVEDKTTTSKETVLDQKSAEKADSVEQQDGALEKTRALGLSESPAEGS----- 1599  
QY 1503 SDVTIPPKNSTNHERK-----STLPTPTQRESIPAK--SPVPG--VDPVVSHPDPPHHRG--- 1553  
Db 1600 -----KAREQEKYKWEQDVVQGHRETSPTRGEPVGGQKEPVPWAEKGSPEQEVRYW 1651  
QY 1554 -----STAGEVYNHSLPTQLDPAHPFHRLDPA---AYLFQRLSPTPGYPSQYLIAM 1606  
Db 1652 RDRDITLQDAYWREL--SCDRKWMFPHELDQGARPRYCEERESTFLDEGPDEQETIPL 1709  
QY 1607 ENT--RQITLNDYITSQOMQVNLRPDVARGLSPREQPLGLP-----YPATRGLIDLTNMP-- 1659  
Db 1710 QHTPRSPWTSDFKDFQELPKQKLEVERWLA--ESPVGLPPEEDKILTRSPFEILISPPAS 1767

QY 1660 -----PTILVPHPGGTSTPPMDRITYIPGTQITFPP-----RPNYSASMSG- 1701  
Db 1768 PPEMTGORVPSAQOESPVPDTESTAPMRNETTPSWLAELIPWVPKDRPLPPAPLSAP 1827  
QY 1702 -----HPTHLAAAASAEERE-----REREKERERERI 1729  
Db 1828 APPTAPEPHTPVVPSWGLAEYDVVAQVGAEELEGGPYSPGLKDYKRAEGEREGG 1887  
QY 1730 AAASSDLYLRPGSEOPCRPGSHGYVVRSPSVRTOETMLQORPSVFOGTNGTSTVITLDP 1789  
Db 1888 AGAPDSSFSFKVPEAGESLATRDETEPEQREPTYPIDERSFYADIYEQMMLTGLGP 1947  
QY 1790 TAOIRIMPLPAG--PSISQGLPASRYNTAADALAAALVDAASAPQMDVSKTKESHEA 1846  
Db 1948 ACTREPPLGASGOWPHLSTKEAAGCNTSAE-----KETSSPA 1987  
QY 1847 ARLEENURSAANVSEOOQLEQKLEVEKRSVQCXYTSSAPPS-----GKPOPHSSVYVS- 1901  
Db 1988 S--PQNLSQDTPAFS-----YASLAGPAVPPRQEPDGPVNEPSI 2025  
QY 1902 -----EACKDKGPP-----PKSRYEEELRTRGKTTITAAINFIDVITROI 1941  
Db 2026 TPAPVPPRAPISLKSGLSPPLNGSTVSCSPDRTPSPKETGRGHWDGTDNDSDL----- 2079  
QY 1942 ASOKDARERSQ-----SSDSSSSLSSSHRY-----ETPSDAI 1973  
Db 2080 --EKGAEREQEKETRSPSPHPMPMGHSSLLWPEATEYSSLSDDLHSGVRSPLDFPASAF 2137  
QY 1974 EVISPASSAPPQEKLYQYEVVVKANOQ-----ENDPTRQYEGPLUHY 2017  
Db 2138 GFSS--LQAPAPPQ-----LPSPAEPRSAPCGSLAFSGDRALALVPGTPTTRH--DEY 2186  
QY 2018 RPOEESPOOQLP--PSSQAEQ---MGQVPRHRLITLADHICQIITQDFARNQVSSOT 2072  
Db 2187 LEVTKAPSLDLSLPQPSPPSPGGLLSNLP----- 2218  
QY 2073 PQOPPTSTFQNSALVSTPVRTKTSNRYSP-----ESQAQSVHHORPG 2116  
Db 2219 --PASPALSEGSSSEATTPIVSSVAERFPGLAEAAQSAEGLSGKESAHSLLDWTPL 2275  
QY 2117 SRV-----SPENLVKSRGSRPKSPERSHVS 2144  
Db 2276 SPAPSASLDLAPAPAPAPAPGLPGLDGTLPCEPCTGELTKKPSFLSPSGDHEAN 2335  
QY 2145 EPYE-PISPPQ--VPVHEKODSLLLLS--ORCAEPAEORNDARSPGSISYLPSTFKLENT 2201  
Db 2336 PGGETSLNPPGFTVATAKEEAEAPHAWERGSWPEGAERSRPD-----TLUSSE 2385  
QY 2202 SPVVKSKQEIFRKLNSGGGDSMAAAQPGTEIFNLPAVTTSGSVSRGHSFA--DPASN 2260  
Db 2386 QPLRPGK-----SSGG-----PPCSLSSEVAGPQGCATDRPH 2419  
QY 2261 LGLEDIIRKALMGSDDKVEDHGVMSOPMGVVPGTANTSVTSGET-----RREGD 2313  
Db 2420 CG-----ELSPSF-----LNPL--PPSTDDSDLSTEEARLAGKGGRRVR 2459  
QY 2314 PPSHSGVC-----KPKLISKNSRKSQKPIPGQ-----GYLGE 2348  
Db 2460 PGA--TGGCPMADETPTPTSASDGSOSDSVPPPETECPSITAEALDDEGDGFLPVD 2518  
QY 2349 RPSVSVSHSEGDYHROTPGWAWEDRPSSTGCTQFPYFNPLTMRMLSTPTPTIAC----- 2403  
Db 2519 KAGVSGCTHHPRCHDPPPTPLDPPPS-----PPRDPVCMADPE 2558  
QY 2404 -----APSAVNQAAPHQONRI--WEREPAP 2426  
Db 2559 GLSSESGRVERLREKGRPRRAFCRAKPAFPARLDIRGRSRPT 2603

RESULT 7

QRMSP1

microtubule-associated protein MAP1B - mouse





|    |      |                |       |    |    |   |    |   |   |   |     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |      |      |      |
|----|------|----------------|-------|----|----|---|----|---|---|---|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|------|------|------|
| Qy | 766  | SLAVPSTKPAEDSV | TQVND | IS | AE | T | AQ | M | D | V | D   | Q | O | E | H | S | A | E | G | S | V | C | D | P | P | P | A | K | A | D | S | V | D | -    | 824  |      |
|    |      | :              |       | :  |    | : | :  | : | : | : | :   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :    |      |      |
| Db | 6358 | PVRVPEY        | -P    | K  | E  | L | E  | P | E | V | A   | F | E | E | V | T | H | V | E | L | V | E | E | E | E | E | E | E | E | E | E | E | E | -    | 6611 |      |
|    |      | :              |       | :  |    | : |    | : |   | : |     | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   |      |      |      |
| Qy | 825  | EVVRPEN        | H     | A  | N  | K | S  | V | E | G | D   | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 878  |      |      |
|    |      | :              |       | :  |    | : |    | : |   | : |     | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   |      |      |      |
| Db | 6612 | PVKVPEV        | P     | R  | K  | P | P  | E | E | K | P   | V | P | V | P | K | K | E | A | P | K | A | P | K | E | K | V | P | V | L | P | K | E | -    | 6671 |      |
|    |      | :              |       | :  |    | : |    | : |   | : |     | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   |      |      |      |
| Qy | 879  | SSATCS         | A     | D  | E  | - | -  | - | - | - | -   | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 914  |      |      |
|    |      | :              |       | :  |    | : |    | : |   | : |     | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   |      |      |      |
| Db | 6672 | VPKKVP         | E     | K  | V  | P | V  | P | V | P | K   | V | E | A | P | P | K | A | P | P | K | V | P | P | A | P | K | K | V | E | A | P | K | -    | 6731 |      |
|    |      | :              |       | :  |    | : |    | : |   | : |     | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   |      |      |      |
| Qy | 915  | LVSSPL         | K     | N  | P  | L | D  | L | P | Q | L   | H | R | A | A | I | P | P | W | - | - | - | - | - | - | - | - | - | - | - | - | - | - | -    | 972  |      |
|    |      | :              |       | :  |    | : |    | : |   | : |     | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   |      |      |      |
| Db | 6732 | LIP            | E     | E  | K  | K | P  | T | - | - | -   | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 6785 |      |      |
|    |      | :              |       | :  |    | : |    | : |   | : |     | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   |      |      |      |
| Qy | 973  | LEQROR         | Q     | E  | I  | D | L  | E | C | R | S   | S | T | S | P | G | T | S | K | N | R | E | V | I | Q | A | P | H | O | L | I | T | N | -    | -    | 1028 |
|    |      | :              |       | :  |    | : |    | : |   | : |     | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   |      |      |      |
| Db | 6786 | PEEEV          | L     | P  | E  | E | E  | V | L | P | E   | E | E | I | P | E | E | E | V | - | - | - | - | - | - | - | - | - | - | - | - | - | - | -    | 6843 |      |
|    |      | :              |       | :  |    | : |    | : |   | : |     | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   |      |      |      |
| Qy | 1029 | PTRP           | P     | P  | P  | L | P  | S | S | T | T   | V | A | S | E | K | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | -    | 1085 |      |
|    |      | :              |       | :  |    | : |    | : |   | : |     | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   | : |   |      |      |      |
| Db | 6844 | PKVP           | P     | A  | P  | P | E  | I | K | K | V</ |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |      |      |      |

|    |      |                                                               |      |
|----|------|---------------------------------------------------------------|------|
| Qy | 1782 | SVITPLDPTAQLRIMPLPAGGSPISIGLFPASRYNTAADALAALVDAASAPOMDVSKTKE  | 1841 |
| Dy | 7422 | PLVVPKKPEAP-----PAKVPEVPKEV-----VPEKKVAVPKK                   | 7454 |
| Qy | 1842 | SKHEAARLEENLRSAAYSEQOOLEOKTLEVEKRVSQCPLYTSAFPSSKGPOPHSSVVYS   | 1901 |
| Dy | 7455 | PVPPAKVPE-----VPKPKVLEE-----PAVPVPE-----                      | 7481 |
| Qy | 1902 | EAGDKGGPPKSYEBELTRCKTTITAANFIDVIITRIQIASDKDAREGQSDDSSSL       | 1961 |
| Dy | 7482 | --RAESPPPPEV-YEE-----PEEIAPEEEI-----APEEEKPV                  | 7511 |
| Qy | 1962 | SSHREYTPSDAIEVISPASSAPPQEKLTQYQEVVKANQA-----ENDPTRQYEGLHHY    | 2017 |
| Dy | 7512 | PVAEEEEPE-----VPPPAVPEEPKKIPEKKVPVIKKPEAPPKEPEKVIKPKLK        | 7566 |
| Qy | 2018 | RPOQESP-----QQOLPPSSQAEQGQVPRTHRLITL-----                     | 2051 |
| Dy | 7567 | RPPPPPPAPPKEDVKEKIFQLKAIPKKKVPENQPQVPEKVELTPLKVGGEKKVRLLPER   | 7636 |
| Qy | 2052 | -ADHICOIITODFARNQVSQTPO-----QPPTSTFTQNSPALYSTVPRT             | 2095 |
| Dy | 7627 | KPEKKEEWLVLSVRKRBEPEEKVEPKKLEKVKKPAVPEPPPP-----KPVEEVEVPTVT   | 7682 |
| Qy | 2096 | KTSNRYSPESAQSVHHQPGGRSVSPENLVDRSGRSKPSERSHVSSEPY-----I        | 2150 |
| Dy | 7683 | KREKIPETPKVPEIKPAIPLAPEP-----KKKPAENVKTIRPPVPEPEPTPIA         | 7732 |
| Qy | 2151 | SPQQVPMVHEKBQDSLLLLSORGAEPAEQNDARSP-GSISYLPSFFTKELENTSPMWXSKK | 2209 |
| Dy | 7733 | APVTVPVWGK-----AAKAPKEEAAPKGPITKGVPK-----KTPSPTEAER-          | 7775 |
| Qy | 2210 | QEIFRKLNSGGGDSDMAAQCTEFNLPAVYTTSGSVSSRGCHSFADPNUGLEDIIRK      | 2269 |
| Dy | 7776 | ----RKLRPGSGGEKPPDEA-FFT--YQLKAVPL-----                       | 7802 |
| Qy | 2270 | ALMGSFDDKYVEDHGVMVSPMG-----VYPGTANTSVYTSGTRREGDPSPHSGGV       | 2321 |
| Dy | 7803 | ----KFVKEIKDIILTESEFEVFGSSAIFECLVSPSTAIIITWMKDGSNIRE---SP---- | 7849 |
| Qy | 2322 | CKPKLISKSNRK                                                  | 2334 |
| Dy | 7850 | -KHRTADGDKRK                                                  | 7861 |

## RESULT 9

T15348  
 hypothetical protein B0350.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #tex  
 C:Accession: T15348  
 R:Gattung, S.  
 submitted to the EMBL Data Library, February 1996  
 A:Description: The sequence of C. elegans cosmid B0350.  
 A:Reference number: Z18332  
 A:Accession: T15348  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-5170 <GAT>  
 A:Cross-references: EMBL:U50071; NID:g1208871; PID:g120  
 C:Genetics:  
 A:Gene: CESP:B0350.1  
 A:Introns: 48/1; 5039/3; 5116/3

|                       |                   |                    |             |              |
|-----------------------|-------------------|--------------------|-------------|--------------|
| Query Match           | 3.5%              | Score 440;         | DB 2;       | Length 5170; |
| Best Local Similarity | 18.7%;            | Pred. No. 4.3e-09; |             |              |
| Matches 556;          | Conservative 357; | Mismatches 1101;   | Indels 958; | Gaps 126;    |

```

Qy 14 STEGSRYPHVSQVTFPNTRHQEF-----AVPDYRSSHLEVSQASQLLQQ 60
 :||| ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 775 SSGCOSEEPHIVKETTNTTTRRELDEPEKGNVTFSPASSHAESEROVPESPVYSHOE 834

```

QY 61 QQQ--OLRRRPSILLSEPH-----PGSDRP-----QERRTSYEPFHP--GPSVPVDHDSLES 106  
Db 835 YPHVVETITTTTINVTNINYYDDNEDNVPSSDEPATQHQSETSVHSHSDPSVVEESDQEGGLGS 894  
QY 107 KRPLEQVSDSHFORVSAVPLVHPILPEGLRASADA--KDPDAPFGKHAPSPSI-----S 161  
Db 895 KVLGFAKKAGMAGVAGVAAVALA---AVGAKAAYDAFEKDDDEDEDSHSPSPVPEYQS 951  
QY 162 GQPCGDONAS-----PSKLSKEELIOSMDRVDREIAKVEQQILKLLKKQOQLEBE 212  
Db 952 EYQODSAQSSHTDFEHMHPESPTEHEKTEEFDH-----SHPESPVLSEKEREHQVTS 1006  
QY 213 AA-----KPPEPEK-----PVSPPP-----VEQKHSRIVQII----- 239  
Db 1007 TITTTTITREYNDEPEDEQHQGHSPAPSSHTAEHPHIVETITTTTITVREFQEEPEEL 1066  
QY 240 ---YDENRKAEEAKIPEGLGPKVELPLYNQPSDTKYVYHNINIKTNQVMRKLILFKRR 296  
Db 1067 EYKQEDNSRKSPPSH-----LDSERLSPAQSPMEV-----SQENLVTEITTTTITVTS---YYDEP 1104  
QY 297 NHARKOREQKICQRYDOLMEAEKKV--DRLENNPRKAKESKTRYYEYKQFPEIRKORE 354  
Db 1105 EHFEQETGKSPAPSSHVES--EQVLSPVADPRH--VMETITTTTITVTRQFHDEDSERS 1162  
QY 355 QOERFORVQORGAGLSATIAARSEHSEIIDG-----LSEQE--NNEKOMRQLSVIP 404  
Db 1163 DSPNRDVV--EAQSIHSSNIESHQFSEKDDDSQSGSPVKSDEPVKHSYKFETSTTE 1220  
QY 405 PMMFDAQORRVKFINMGLMB---DPMKYKDRQFMNVTWTHEKEIKFKFIQHPKNF-- 459  
Db 1221 TREFDRPE-----LDSERLSPAQSPMEV-----SPITTEENIVRPSVSK 1261  
QY 460 -----GLIASYLERKSVPCDVLVYLLTKKNYKALVRRNVRKRRGNQOAIARPSOEK 513  
Db 1262 SEYSGEHPVSVIETKTTTITVREFYDDQDE-----HENQTSQEEELASSIPTEEEE 1313  
QY 514 VEK--EEDKAEKTEKE--EKKDEEKDEKEDSKENTKEKDIDGTABETEERQAT 568  
Db 1314 HEGSHLFKETTITTTTITVREFYDEPENVEELQDPQSPAPSSHVESGHASESPVAQOQEI 1373  
QY 569 PRGKTANSQORRGRITRSMTNAAAA-----SAAAAATEEPPLPPPPPISTEP 622  
Db 1374 PQ-----TREFHSPAQAQYFHEEYESHVLTEQAPLLTEQOHPESGDE 1418  
QY 623 VETSRWTEEMEVAKK--GLVHGGRNMAIA--KMVGTSKSEAQCKNFYFNKRRHNDLNLQ 680  
Db 1419 SDGEGLSKVLGFAKKAGMAGVAGVAAVALAAGAKAAYDALKKDDDEDEQERESILR 1478  
QY 681 QHKO-----KTSRRPREERDVQCESVASTVSAQDE-----DIEA 716  
Db 1479 QERSIDSPHASEQSIIEEHERFEESPVPSEKHHVTITTTTITVREFYDEHEPLVSQEI 1538  
QY 717 SNEENPDSE-----VEAVKPSDESPENAT--SRGNTEPAVE 752  
Db 1539 EGEKSPADSEKSLPHVVETITTTTITVREFDKNDSESPVSEKQEPTTVSREYVETAE 1598  
QY 753 LEPT---TETAPT-----SPSLAVPSTKPA-----EDRSVE 781  
Db 1599 DEPEHHYETITTTITVKEVIDDSQEMGDDDEPKQESQVQVETITTTTITVREYDNDDETR 1658  
QY 782 TQVNDISAETAQOMDV-----DQEHSAEGSVCDPPP-----ATKADSDV 824  
Db 1659 SEAGDSHITETTTTITVREFPHGQPEETEEDVEELPPKIEEDNVSEYSESSTSVSR 1718  
QY 825 EVRYPENHASKVEGDNKEDLDRASEKVEPRDEDLVVAQOINAQRPQSDNDSSATCS 884  
Db 1719 EVRDEPHIETITTTTITVREYHNEPEETYDDQD---AAPISFS--QEHODDSQAS-- 1771  
QY 885 ADEDVDCGEPERQMPMDKPSLLNPTGSIILVSSPLK-----PNPLDLPQLQH 932  
Db 1772 ---HDQDRSPVESEKSVKHTTITTTTITVTRQLYDDEASEINCESPVATEEH 1824  
QY 933 RAAVIPMVSCTPCNIPITGTPVSGVALYVORHIKAMHESALL-----EQQRQOE 982

Db 1825 -----VSSTKSDDE-----SEQHVPSVIEITTTTITVREFYDQDDELQREDH 1866  
QY 983 IDLECRSTSPCGTSKSPNREWEVLQAPAHQOLITNLPEGVRLPTTRPTR---PPPLIPS 1039  
Db 1867 TQSERRSSIP--TEETEHEDSHLIKET-----TTTTTITVREFYDEPENVEK 1911  
QY 1040 SKTTVASEKPSFIMGSSISQCTPGTYLTSHNOAS--YTQETPKPSVGSISLGLPQOQESAK 1098  
Db 1912 LQDSQFSLSPS-----SHVESIIVPESP-----VAKQOEIPQ 1944  
QY 1099 -----SATLPYIKQEFSPRSONSQFGLLVRAQHEGVVGRGTAGAIQEGSITRGTPS 1151  
Db 1945 TREFHEDSPAAYPHEDYEYHQVTEQAP--LLTEQHQHP-----ESGEESDGEFG 1994  
QY 1152 KISVESTPSLRGSIQTQTPALPQTGIPTEAL--VKGSIISRMPIEDSSPEKGREEAASGHV 1210  
Db 1995 S--KVLGFAKKAGMAGVAGVAA---VALAAGAKAAYDALKKEDD--EEDQEEERS--L 2045  
QY 1211 IYEGKSGHILSYDNIKNAREGTRSPRTAHEISLRSVESVEGNIKQGMMSRSPVSAPE 1270  
Db 2046 IREER-----SFD-----SPHASEQSOIEKEHKEFE-----ESPVPSEKH 2080  
QY 1271 GLICRALPRGSPHSDLKE--RTVLSSGSIQGTTPRATTESFDGLKYPKQIKRESPPIRAF 1328  
Db 2081 HDQSSALPQESVSQPIEKESRTFNDESEFGVYKSDHYTDDQESLSKPE----- 2129  
QY 1329 EGATYKGPYDGIITIKEMGRS---IHEIPRODILTQESRKTPEVQOSTRIIE----- 1379  
Db 2130 -----SGEAFSQFTSEKQDRSDPIHS--QKEDISQFQNESSPEDVVKSEQPHDEKPDLE 2183  
QY 1380 --GSGISQGTPIKFNNGSGSAIKHNKSLITG--PSKLSRGMPPLEIVPENI----- 1427  
Db 2184 ROGYSYSGSPKSPGGS-----ITGLDEERKALSGVQEPEDRPFNAESHEKTE 2231  
QY 1428 -----KVVERGKYEDVAGETVRSRHTS--VWSSGSPVLSRTLHEAPKAQLSPGIYDTS 1480  
Db 2232 ATSDENLFESDKY---APASVPSPEDSSNRVIETTTTITVTRHEFEPE-----DDHS 2280  
QY 1481 ARRTPVSYQNTMRSGSPMMNRTSDVTIPPKNKSTNHERKSTLTPT--QRESIPAKSPVPCVD 1539  
Db 2281 ---YVSESQYSSSGSP-----VPEKSDVRVIETTTTITVTRHEFEADEIPTI-- 2327  
QY 1540 PVWSHSPDFPHRGSTAGEVYWSHLPTQLDPAMPFHRALDPAAAAAYLFQRLSPTPGYPS 1599  
Db 2328 -----VSSH-----DDQAA-----SPVSPSEED 2345  
QY 1600 QYQLYAMENTROTIL-----NDYITSQOMOVNLRDPVARGLSPREQPLGLPYPATRGII 1653  
Db 2346 VHEQIQTTTTTTTITVTRKEHFPDDEIDSEHMN-----ESDKYASGSPVSPSEEDSS 2394  
QY 1654 DLTNMPPTILVP---HPGCTSTPPMDRITVPGTQITFPFPPRPNYSASMSGPHTHLAAA 1709  
Db 2395 RVETTTTTTITVTRHEFPEDDHSVPVQTOEY-----SASESP-----VPSE 2435  
QY 1710 ASABER-----ERERERERERERERERERERERERERERERERERERERERERER 1745  
Db 2436 KSERVIETTTTTTITVTRHEFDEHDHILGQOQESDQIPSEISITSENMD---RETSSSPVQ 2492  
QY 1746 -----GRPGSHGYVRSPPSVRT-----QETMLQORPSPVFOGTN 1779  
Db 2493 SNRDEEVLPAIAPYKQPTQYGRVDSHDAPASPAESPISQAYKQEEQSAHSLFOFQ 2552  
QY 1780 GTSVITPLDPTAQL-----RIMPLPAG--GPSISQ----- 1807  
Db 2553 RSSVSHEDSPAAYFHDNEDHSDSPVSDRPLLTQEQHPESGEESDGEFGSKVLGFA 2612  
QY 1808 ---GLPASRYNTAADALAALVDAASAAPQMDVSKTESKHEAARLEENLRSRSAVSEQ 1864  
Db 2613 KKAGMAGVAGVAAVALAAGAKAA---YDALKKDDDEDENPDQEKLLPKS---PERQ 2665  
QY 1865 QL-----BQKTLVEKRSVQCLYTSSAFPSGKPPQPHSSVYVSEAGDKGPPPKSYEEELR 1920



Db 2666 VLVNVPERSDSEISE--IELEYTS-----PSPTEKSESCY 2699  
Qy 1921 TRGKTTITAAFNIDVITITQASDKDARERGSSQSSSSSSSHRYETPPSDAIEVISPAS 1980  
Db 2700 TETVTTTT-----VTREYLD-----PQSVTRSRSPSEH-----DISQYAPES 2738  
Qy 1981 -----SPAPQOE-KLQTYQBEVVKANAQENDPTRYEGPLHHY----- 2017  
Db 2739 PVEQDPYVVEKTTTIRQYHDEPPQIEEQIPBEVTVLREVYESP--EGDEPEQHYET 2796  
Qy 2018 -----RPOQESPS----- 2025  
Db 2797 KTTTTIKEVHVVEDVQISPVHSERSVSEKQLPABQLDEPVTETTTATVTRERYEE 2856  
Qy 2026 PQOQLPPS--SOAEGMGQVPRTHRLITLADHICQI--ITQDFARNQVSSQTPOOPPTSTFQN 2083  
Db 2857 PEEVPPSGSEADDESHAPKYMETTTTTTITREYEVSEDEDEHQSQSVORDESPAPSEDS 2916  
Qy 2084 SPSALVSTPVTKTSNRYSPESQAQSV-----HHQRPGRSVSP 2122  
Db 2917 VKHVIEKTTTTTVEERYEPEDSHSPVPSDEDDVHGFVKTTTTTTTTTTHHEPPEDEPPSDE 2976  
Qy 2123 NLVDKSGSRPGKSPERSHVSSEYEPISPPQVPVH--EKODSLLLLSQRGAEPAEQRND 2181  
Db 2977 HVESERTAGSPVSEED--SSRIETTTTTTTRHFELED-----DQEHVSESQYSA 3030  
Qy 2182 ARSPGISYILPS-----FFTKLENTSPMKVKKQETFRKLNSGGSDSDMAAAQPGTE 2234  
Db 3031 SGSP-----VPSEKSERVETTTTTNTRHFEHEDDIPTIVTSHDDPAASSVPSEED 3085  
Qy 2235 IFNLPAVTTSGVSSRGHSFADPNASNLGLEDIIIRKALMGSDDDKVEDHGVVMSQPMGVYP 2294  
Db 3086 VHGOIQTTTTTTTRHVVPDEIDSGRDELEKYSSESVPVSEEDSSRVIE----- 3138  
Qy 2295 GTANTSVTSETREGDPSHGGVCKPKLISKNSRKSIPGQ----- 2342  
Db 3139 -TTTTTTS-----REHFEEDHSHVSESQYASG-----SPVPSEKSERVETTTTT 3187  
Qy 2343 -----GVLGTER--PSSVSSVHSEG--DVH--RQTPGWAMEDR----- 2374  
Db 3188 TTVTRHFEDEEDYIPESRTSHDDGITDQHVPSQSPVSEEDDHVTHEQIIRDDPIDEQ 3247  
Qy 2375 -----PSSGTGTFQFYNPLTMRSLSTPTPTACASVNAQAAPHQONRI-- 2419  
Db 3248 IVESHEYKSESSIPSEQDSTH-----VIETVMSTPTITSERYDPEVEKDVVESADDEIDS 3301  
Qy 2420 -----WEREPAPLLSAQYETLSDSD 2440  
Db 3302 STAOYKSESPVQTEKSLLLAKQOQESGDESD 3333

RESULT 10  
I49505  
adenomatous polyposis coli protein - mouse  
N:Alternate names: APC  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 13-Aug-1999  
C:Accession: I49505  
R:Su, L.  
Science 256, 668-670, 1992  
A:Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the  
A:Reference number: I49505; MUID:92263101  
A:Accession: I49505  
A:Molecule type: mRNA  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Residues: 1-2845 <RES>  
A:Cross-references: GB:M88127; NID:g191991; PIDN:AAB59632.1; PID:g191992  
C:Superfamily: adenomatous polyposis coli protein

Query Match 3.58; Score 437; DB 2; Length 2845;  
Best Local Similarity 19.4%; Pred. No. 2.7e-09;  
Matches 423; Conservative 471; Mismatches 831; Indels 658; Gaps 91;

Qy 347 PEIRKQEQOE--RFQVRGQRGAGLSATIARSEH-----EISEIIDGL 387  
Db 979 PSVESYEDDESCKSCSYQYADLAHKLHSANHHDDNDGELDTPLNYSLKYSDEGLNSGR 1038  
Qy 388 SEQENNEKQKQQLSVIPPMFDAQRRVKFINMGLMBDMKVKYKDRQFMNVTDEKEI 447  
Db 1039 QSPSONERWARPKHVEIDEIKQNEQARQSNTS-----YPTVISEN-----TD----- 1081  
Qy 448 FKDFIOHPKNGFLIASYLERKSPDCVLYY-----VLTAKNENYKALV 491  
Db 1082 --DKHLKFPQPHFG-----QOECVSPYRSRGTSSETNRMGSSHAINONVNQSLCQ 1129  
Qy 492 RRYTKRRGRNQIARPSQEEKVEEKDKAEKTEKKEEKKKEEKE----- 540  
Db 1130 EDDYEDDKPTNYS--ERYSEEQHEEERPTNYSIKYNEKHHDQPIDYSIKYATDISS 1188  
Qy 541 -----KEDSKENTKEKIDGTAETEERQATPRGRKKTANSQGRRKGRITR----- 587  
Db 1189 SQKPSFSKNSAQSTK--PEHLSPSENTAVPPSNAKRQNLRPSSAQORNQTKGTTC 1247  
Qy 588 -----SMTNEAAAAA----- 602  
Db 1248 KVPINQETIQYCVEDTPICFSRCSLSLSSADDEITGCDOTTQEADSANTLOAEVKE 1307  
Qy 603 -----ATEEPPPPPLPPPEPISTEPVETSRWTEEMEYAKKGL--VEHGRNNAALAKMVG 656  
Db 1308 NDVTRSAEDPATEVPAYSONARAKP-----SRLOASGLSESTRHNKAVEFSSGA 1357  
Qy 657 KSEAOCKNFYNYKRRHNLNLLQHKQKTSRKPREE-----RDVQCESVASTVSAQE 710  
Db 1358 KSPS-----KSGAQTPKSPPEHYVQETPLVFSRCTSV--SSLDSFE 1396  
Qy 711 DEDTEASNEENPEDSEVAEVPSE--DSPENA--TSRGNTPEPAVELEPTTETAPSTSPS 766  
Db 1397 SRSTASSVQSPCGSMVSGIISDLPSQGTMPPSRSKTPPP--PPQTVQAKREYVK 1453  
Qy 767 LAVPSTPAEDESVEVOTVNDISAEAEQMDVDQOESAE--GVC-----DP 813  
Db 1454 SKVPAAREKREGPKOTAVNAVO--RVQVLPDVTLLHFAFESTPDGFCSSLSALSIDE 1512  
Qy 814 PPATKADSDVDEVRY--PENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQAINAQR 871  
Db 1513 PFIQK-----DVELRIMPVQ-----ENDNGNETESEQEESNEQDKREV-----EKP 1555  
Qy 872 EQQSD--NDSSATCSAEDVD-----GEPEQRMPMDSKPSSLNPTCSILVSSPLK 921  
Db 1556 DSEKDLDDSD-----DDDIIEECIISAMPTKS-----SRKAKKLAQTASKL--PPVPA 1604  
Qy 922 PNPLDLP-----QLQHRAAVIP-----PMVSC-----TPCNIPIGTVPVSYAL----- 959  
Db 1605 RKPSQLPVYKLLPAQNRLQAKQHVSTPGDDVPVRYCVGPTINFSTATSLDITISPP 1664  
Qy 960 -----YORHIKAMHESALLEEQROREQIDLE 986  
Db 1665 NELATDGVRAIGSGEFERDRTIPTEGRSTDDAQRGKISSIVTDPDLDNKAEEGDILAE 1724  
Qy 987 CRSSTSCGTGSKSPNREWEVLQAPAHQILITNLPGVRLPTTRTPRPPPLIPSSKTTVAS 1046  
Db 1725 CINSAMPKGSKHKPRYKKTMDVQOQASSTSSGANKNOVDTKKKKTPSPVKPMQNTYR 1784  
Qy 1047 EKPSFIMGSGSQCTPGCTYLTSHNQASVTOETPKPV-----GSIS 1087  
Db 1785 TRVRKNTDSKVNVTETTF-----SDNKDSKPSLQTNKAFNEKLPPNEDVRGTGA 1837  
Qy 1088 LGLPROQESAKSATPLYIKOEESFSPRSQSOPEGLLYRAQHEGVYRGTAQIESIT-R 1146  
Db 1838 LDSPHHYTPIEGT--PYCFSRNDSLSLDDDDVDLSREKAEKRGKESKDEAKVTCR 1895  
Qy 1147 GTPTSKLSVES-----IPSLRGS---ITQGTALPQT-----GIPT--EALVKSGSIS 1188  
Db 1896 PEPNSSOAAASKSQASIKHPANRAQSKPVLOKQPTFFQSSKDGDPDRAAGTDEKLOLAIE 1955





Qy 168 -DONASPKLSKEELIQSDVRDE-----IAKVQOILKLLKKQOQ----- 208  
Db 266 LQKRVSISR-SQHDILQMICSKRRSGASEANLIVAKSWADVWKLGAQOTQVKIKHGQOR 324  
Qy 209 -LEEEAAKPPPEKPV-----SPPV-----EQKH-----RSIVQIYD 241  
Db 325 SMNKRORRPAKPKPVGVEGSHQFSTGHANSPTCTIIGKAHTEKVHVPARPYRVLNNFISN 384  
Qy 242 ENRKAABEAHIFBGLGPKVELPLYNOPSDTKVYHENIKTNQVMRKKLI----- 290  
Db 395 QMDFKED-----LSGIAEMFTPVKEQQLSTCHIASNSNLLGQFOQOTDSEEPPLL 440  
Qy 291 -----LEFKRRNHAKQREKICORYDQLMKAEKKVDRIENNRKAK-----ESK 337  
Db 441 PSESFGGNVFFSAQ--AAQPSDK-CSASPLLRQOCIRENGNVAKTPTNTYKMTSLETK 498  
Qy 338 TREYYEKOFPEIRKQOEORQORQVQAGLATSATIARSEHIEISIDGLSEQENNE----- 394  
Db 499 TSD-----TETPSKTVSTVNRSG--RSTEFNRIOQLP--VESKSEETNTEIVE 543  
Qy 395 ---KOMROLVIPPMFADAEORRVKEFINNGLM---EDPMKVYKDRQPMVWTDHEKEIF 448  
Db 544 CILKRGQKATLL-----QORR-----EGEMKEIERPFETKE-----NIELKENDE-- 584  
Qy 449 KDKFTQHPKNFGL-----IASYLERKSVPCVLYLYLTTRKKENYKALVRRNYGKRGRNQ 504  
Db 595 KMKAMKRSRTWGOKCAPMSDLTDLKSLPDTLM-----KOTARQON 625  
Qy 505 IARPSQEBKVEKEBDKAEK-----TEKKEBKDEEKEDEKED----- 543  
Db 626 LLQOTDHAKAPKSEKGTIKMKPCOSLOPEPINTPTHTKQQLKASLGKVGKVEELLAVGKF 685  
Qy 544 -----SKENTKEKKIDCTAEET---ERREQATPRGKTANSOGR----- 581  
Db 686 TRTSGETHTHREPAGDGCSIRTFKESKQIILDPAAVTRGMKMPRTPKBAQSLDLIAG 745  
Qy 582 -----KGRITRMTNEAAAAAATAEPPPLPPPEPTEPETSRSW----- 628  
Db 746 FKELFQTPGPEESWTDEKTKIACKS-----PPPESDVT-PTSKQMPKRSR 793  
Qy 629 ---TEEMEVAKKGLVEHGRNWAIAKMWGKSEQAQKNFYNTKRRNLONLLOQHOK 685  
Db 794 KADVEEFELALRKLTPPSAGK--AMLTTPAGGDEKIDKAFMGTPVQKLDLAGTLPGSRQ 851  
Qy 686 TSRKPREE---ROVSQCESVASTVSAQEDEDIAASNEE---ENPEDSEVAVKPSD 736  
Db 852 L-OTPKRAQALEDLAGKELFQTPG--HTEELVAAGKTTKIPCDSPQSDPVDPTSTK 908  
Qy 737 SPENATRGNTPEPAVELEFTTETAFSTPSLAVPSTKPADESVEVTONDSI----- 788  
Db 909 RPKRSIRKADVEG--ELLACRLMPSAGKAMHTPKPSVGEEDKIIIFVGTVPQKLDLFTEN 966  
Qy 789 -----SAETAQOM-----DVDOQEHSAEGSV-----CDPPPAKADSV 823  
Db 967 LTGSKRRRQTPKEAQALEDLTFKELFQTPGHTTEAIVAAGKTTKMPCESSPPESADPT 1026  
Qy 824 VEVRYPE-----NHASKVEGDNTERDLDR--ASEKVEPR 856  
Db 1027 STRQPKTPLEKRVQKELSALKLKTQTSGETTHTHDVKPGGEDKSINAFRETAQKLDL- 1085  
Qy 857 DEDLVAAQOINAOPEQSDNDSSATCSADEVDGEPEPRQMFPMDSKPSLLNPTGSTIV 916  
Db 1086 -----AASVTGSKRHPKTKAQAQL-----EDLAGKELFQTPVCTDKPTTHEKTKIAC 1135  
Qy 917 SSPLKPNLDLPOLQH-----RAAVTPPMVSCPTCPNIP 949  
Db 1136 RS--QPDVDPDPTSSKPSQSKSLRKVDVEEFALRKRTPSAGKAMHTPKPAVSGERNIY 1193  
Qy 950 --IGTPVSGYALQ-----RHIAKHAFESAL-----LEORQOEQID 984  
Db 1194 AFMGTPVQKLDLTENLTGSKRRLQTPKEAQALEDLQAGKELFQTRGHTESMTNDKTA 1253  
Qy 985 LECRSS-----TSPCGTSKS 999

Db 1254 VACKSSQDLDKNPASSKRRRLKTSIGKVGKVEELLAVGKLTQTSGETTHTHTEPTGDGKS 1313  
Qy 1000 PNREWEVLQAPHQLI---TNLPEGVRLPTTRPRPPPLIPSSKTTIVASKEPFSIMGGS 1056  
Db 1314 ---MKAFMESPKOILDSAAASLTGSKRQLRT-----PKGSEVPEDLAGFI---E 1356  
Qy 1057 ISOCTPGYLTSHNOASVTOE--TPKPS-----VGSISLGLPROQESAKSATLPYIKO 1107  
Db 1357 LFO-TP-----SHTKESMTNEKTTKVSYASQDPLVDTPTSKQPKKSLKAD-----TE 1406  
Qy 1108 EEFSPRSQNSOPEGLLVRQAHEGVVVRGTAGAIQEGSITRGTPTSKISVPSLRGSITQ 1167  
Db 1407 EEFLAFRKQTPSAG--KAMH--TPKPAVGEKDKINTFLGTVPVK----- 1446  
Qy 1168 GTPALPQTG-IPTREALVKGISIRMPIEDSSPEKGREAAKSHVYIEGKSHILLSYDNIK 1226  
Db 1447 ---LDQPGNLP-----GSNRL-----QTRKEKAQ-----ALEELT 1474  
Qy 1227 NAREGTRSPRTAHEISLKRYSVESVEGNIKOGMSMRSPVSAPLEGLICRALPRGS---PH 1283  
Db 1475 GFRELFPQTPCTDNPTADEKTTKKI-----LCKSPQSDPAD-----TPTNTKORPK 1519  
Qy 1284 SDLKERTVLSGSIQGTTPRATTESFEDGLKYPKQIKRESPPIRAFEGAIT----- 1333  
Db 1520 RSLKADVEEFL---AFPKLTPSAGKAMHTPKAAVGEKDKINTFVGTPVEKLDLLGNLP 1576  
Qy 1334 -----KGPYDGIITTIKEMGRS---IHEIPRODILTQESRKTPEVVQSTRPIIE 1379  
Db 1577 GSKRRPQPKKALEDLQAGKELFQTPGHTESMTDDKITEVSKSP----- 1625  
Qy 1380 GSIQGTPIKEDNNSGQ---SAIKHNKSLITGPKSLSRGMPLEIVPENIKVVVERGY 1435  
Db 1626 ---QPDVPKTPSSKQRLKISLGKVGKVEVLPGKLTQ----- 1661  
Qy 1436 EDVKAQETVRSRHTSVSSGSPSVLRSTLHEAPKAQLSPGIYDDTSAR---RTPVSVQNTMS 1493  
Db 1662 ---TSQKTTQT-HRETAGDGKSI--KAFKESAKQMLDPANYGTGMERWPRTPKEAQSLE 1715  
Qy 1494 ---RSPMMNRISDVTIPPNKSTNHERKSTLTPQRESIPAKSPVP-GYDVPVS--HSFPD 1548  
Db 1716 DLQAGKELFQTPD-----HTEEST-TDDKTTKACKSPPPESMDTPTSTRRRT 1764  
Qy 1549 PHHRGSTAGEV-----YWSHLPTQLDPAMPFHR-----ALDPAALYLFQORQLSP 1593  
Db 1765 PLGRDIVEELSALKQLTQTHTDKVPQDEDKGINVFRETAKQLDPAASVTGSKRQ-PR 1823  
Qy 1594 TPGYPSQ--YOLYAMENTROT-ILNDYITSOOMQVNLRPDVARGLSPREQLGLPY----- 1646  
Db 1824 TPKGKAQPLEDLQAGKELFQTPVCTDKPTTHEKTKI---ACR--SQQDPVPVGTPIFKP 1878  
Qy 1647 PATRCI-----IDLTNMPPTI---LVPHPGGTSTPPMDRITYIPGQITPPPRPY 1693  
Db 1879 QSKRSRLKADVEESLALURKRTPSVGKAMDTPKPAGGDEKDKAMPGTPVQKLDLP----- 1934  
Qy 1694 NSASMPGHPHTLAAAAAERERERERERERERERERERERERERERERERERERERERER 1753  
Db 1935 ---GNLPG-----SKRMPQTPKEAQALEDL-AGKELFQTPCTDPTTDEKTTK 1980  
Qy 1754 V--RSPSPSVRTQETMLQORPSVFGTNGTSVITPLDPTAQLRIM--PLPAGGSPISQGL 1809  
Db 1981 IACKSPQDPDVTTPASTKORPK-----RNLKADVEEELFALRKRTPSAGKAMDPK 2032  
Qy 1810 PA---SRYNTAADALALVDAAASAPQMDYSKYTESKHEARLEENLRS-----RSAAV 1860  
Db 2033 PAVSDEKNNITFVETPVQKLDLLGNLP--GSKROPQTPKEAQALEDLVGKELFQTPGH 2090  
Qy 1861 SEQOOLEQKTELEKRSVQCLYTTSSAFSPGKQPHS-SWVYSEAGKDKGPPPKRYEE- 1918  
Db 2091 TEESMTDDKITEVSKS-----POPEFSTSSRSKORLKLPLVKVDMKEEP 2136  
Qy 1919 -----LRTROKTTITANFIDVITROIASDKDAREKSGOSSSSSSLSHR----- 1965

Db 2137 LAVSKLRTSGTQTHTE-----PTGDSKSIKAFKESPKQILDPAASVTSRRQLRTRK 2191  
Qy 1966 -----YETPDSIAEIVS-----PASSAP-----PQE 1987  
Db 2192 EKARALELDVFKELFSAPGHTESMTDKTKIPCKSPPELTDATATSKRCPKTRPRK 2251  
Qy 1988 KLOTYQPVVANKAENDPTQYEGPL-----HHYRPOQESPPQOOLPPS 2033  
Db 2252 EVKEELSAVERLTQTSQSTHTRKEPASGDEGIKVLKQRAKKKPNVBEEPSRRPRAPK 2311  
Qy 2034 SOAEGMGQVPRTHRLITLADHICOLIITODFARNVSSQTPQOPTST----- 2080  
Db 2312 EKAQPLEDLAGTELSETSGHTQSLRAGKATKIPCESPPLEVVDTTASTRHRLTRVQK 2371  
Qy 2081 --FQNSPVALYSTVTRKTSNR--YSPESQAQSVHHQRGSRVSPENLVKSGSRPKQSP 2137  
Db 2372 VQKREPSAVKFTQTSGETTDADKEPAGEDKGIKALKESAKQTPAPASVT--GSR--RRP 2428  
Qy 2138 ERSVSEPEPIPPQVPPVHVKQDLSLLLSQSGAEPAEORNDARSQSGISYLPSPFTK 2197  
Db 2429 RAPRESAQAIEDLAGFKDPAAGHTEESM-----TDDTKTKIPCKSS-----PE 2471  
Qy 2198 LENTS-----PMVKSQKOEIIFRLNLSG-----GGDSMAAOP-----GTEIFNLPA- 2240  
Db 2472 LEDTATSKRRPRRAQKVEKVELLAVGKLTQTSGETTHDKEPVGEGKTKAFKQPAK 2531  
Qy 2241 --VTTSGSVSRGHSFADPASNGLIEDIIRKALMGSDDKVEDHGVVMSQPMGVVPGTAN 2298  
Db 2532 RNDAEDVIGSRQPRAPKKAQPLED-----LASFOE-----LSQTPGHTTELAN 2577  
Qy 2299 --TSVTSGETREEDPPSPHSGVCK-PKLISKSNRSKSKSIPGGYGLGTERP 2350  
Db 2578 GAADSFTSAPKQTPDSGKPLKIRRLRAPKVEPVGVVTRDPVKQSKSNTSLP 2633  
RESULT 13  
S37431  
N:Alternate names: ankyrin long splice form - human  
N:Contains: ankyrin B, 440K splice form; ankyrin-B; brain ankyrin; non-erythroid  
N:Species: Homo sapiens (man)  
C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 13-Aug-1999  
C:Accession: S37431; A35643; B39643; A40334; A49462; S14533; S14569  
R:Chan, W.  
submitted to the EMBL Data Library, September 1993  
A:Reference number: S37431  
A:Accession: S37431  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-3924 <CHA>  
A:Cross-references: EMBL:Z26634; NID:9406287; PIDN:CAA81387.1; PID:9406288  
R:Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.  
J. Cell Biol. 114, 241-253, 1991  
A:Title: Isolation and characterization of cDNAs encoding human brain ankyrins reveal a  
A:Reference number: A39643; MUID:91302466  
A:Accession: A39643  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2077 <OTL>  
A:Cross-references: GB:X56957  
A:Accession: B39643  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1443,3585-3924 <OTT>  
A:Cross-references: EMBL:X56958  
R:Tse, W.T.; Menninger, J.C.; Yang-Feng, T.L.; Francke, U.; Sahr, K.E.; Lux, S.E.; Ward,  
Genomics 10, 858-866, 1991  
A:Title: Isolation and chromosomal localization of a novel nonerythroid ankyrin gene.  
A:Reference number: A40334; MUID:92009921  
A:Accession: A40334  
A:Molecule type: DNA  
A:Residues: 463-474, 'PE', 477-495 <TSE>  
A:Cross-references: GB:N37123; NID:g178647; PIDN:AAA62828.1; PID:g178648

R:Chan, W.; Kordeli, E.; Bennett, V.  
J. Cell Biol. 123, 1463-1473, 1993  
A:Title: 440-KD ankyrinB: structure of the major developmentally regulated domain and  
A:Reference number: A49462; MUID:94075409  
A:Accession: A49462  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-3924 <RES>  
A:Cross-references: EMBL:Z26634; NID:9406287; PIDN:CAA81387.1; PID:9406288  
C:Genetics:  
A:Gene: GDB:ANK2  
A:Cross-references: GDB:127607; OMIM:106410  
A:Map position: 4q25-4q27  
C:Superfamily: ankyrin; ankyrin repeat homology  
C:Keywords: alternative splicing  
F:2-3924/Product: ankyrin 2, long form #status predicted <MAT>  
F:2-1443,3585-3924/Product: ankyrin 2, short form #status predicted <MA2>  
F:63-95/Domain: ankyrin repeat homology <AN01>  
F:96-128/Domain: ankyrin repeat homology <AN02>  
F:129-161/Domain: ankyrin repeat homology <AN03>  
F:162-190/Domain: ankyrin repeat homology <AN04>  
F:191-223/Domain: ankyrin repeat homology <AN05>  
F:232-264/Domain: ankyrin repeat homology <AN06>  
F:265-297/Domain: ankyrin repeat homology <AN07>  
F:298-330/Domain: ankyrin repeat homology <AN08>  
F:331-363/Domain: ankyrin repeat homology <AN09>  
F:364-396/Domain: ankyrin repeat homology <AN10>  
F:397-429/Domain: ankyrin repeat homology <AN11>  
F:430-462/Domain: ankyrin repeat homology <AN12>  
F:463-495/Domain: ankyrin repeat homology <AN13>  
F:496-528/Domain: ankyrin repeat homology <AN14>  
F:529-561/Domain: ankyrin repeat homology <AN15>  
F:562-594/Domain: ankyrin repeat homology <AN16>  
F:595-627/Domain: ankyrin repeat homology <AN17>  
F:628-660/Domain: ankyrin repeat homology <AN18>  
F:661-693/Domain: ankyrin repeat homology <AN19>  
F:694-726/Domain: ankyrin repeat homology <AN20>  
F:727-759/Domain: ankyrin repeat homology <AN21>  
F:760-792/Domain: ankyrin repeat homology <AN22>  
F:793-825/Domain: ankyrin repeat homology <AN23>  
Query Match 3.3%; Score 411.5; DB 2; Length 3924;  
Best Local Similarity 18.7%; Pred. No. 3.9e-08;  
Matches 493; Conservative 357; Mismatches 946; Indels 845; Gaps 114;  
Qy 57 LQOQQOQLRRRPSILSEFHPGSDRQERRTSYBPFHPGSPVDHDSLESKR-----P 109  
Db 1740 LQRRVEQKGRSKLPTRVKGKEDVP--KKTHRP--HPAASP-----LSSSERHAPGSPSP 1792  
Qy 110 RLEQVSDSHFORVSAAVLPLVHPLPEGLRASADAKDPAGFGKHEAPS--SPISGQPCGDD 168  
Db 1793 KTERHST-----LSSSAKTERHPVSPSSKTEKHSVPS--PSAKT 1830  
Qy 169 QNASP-----SKLSKEELIQSDRVDRVETAKVEQIILKKKQQLLEEAAPPEPEKPV 224  
Db 1831 ERHSPASSSSKTEKHSVSPSTKTERH-----SPVSTKTERHPVPS 1872  
Qy 225 P-----PPVEQKHRSIVQIYDENRKKAEAKHIFBGLGKVELPLYNPSDTPKYVHE 277  
Db 1873 PSGTKDRKPPVSPSGRT-----EKHPPVSPGTERKRLPV--SPSGRTDKHQ 1916  
Qy 278 NIKTNQVMRKKLILFFKRRNHAR-----KOREKICORYDQLMBAWKVKVDRIEN- 327  
Db 1917 PVSTAGTKTERHLPVSPSGKTEKQPPVSPSTKTERIETMSVYR--ELMKAFSGQDPSKHK 1974  
Qy 328 -----NPRKAKESKREYKQFEIRKQEQO-----ERFQVY-----GQSG 366  
Db 1975 TGLFEHSAKQKQPKQKRVKVEKEKGPILTQREAAQTENOTIKRGQLPLVTGTAESKRG 2034  
Qy 367 AGLSATATARSE-----HEISEIIDLQSLQENNEKOMRQLSVIPPMFADQRRV 415  
Db 2035 VRVSSIGVKEDAAAGGKEKVLKSHKIPVQSVPEESHR-----ESEPKE 2080



Db 3768 VSTPAEEKLYQTP-----TSSRGSGPIIQPEEPSEHRESSPRKTSILVIVESADN 3821  
Qy 2410 Q 2410  
Db 3822 Q 3822  
RESULT 14  
T25752  
hypothetical protein F45E4.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25752  
R:Wilson, R.  
submitted to the EMBL Data Library, September 1996  
A:Description: The sequence of C. elegans cosmid F45E4.  
A:Reference number: Z20082  
A:Accession: T25752  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2361 <WIL>  
A:Cross-references: EMBL:U70852; PIDN:AAB09135.1; GSPDB:GN00022; CESP:F45E4.4  
A:Experimental source: strain Bristol N2; clone F45E4  
C:Genetics:  
A:Gene: CESP:F45E4.4  
A:Map position: 4  
A:Introns: 60/2; 111/2; 939/3; 977/2; 998/1; 1021/1; 1125/3; 1140/3; 1157/2; 1173/2; 131/2

Query Match 3.2%; Score 409.5; DB 2; Length 2361;  
Best Local Similarity 18.6%; Pred. No. 2.5e-08;  
Matches 523; Conservative 374; Mismatches 975; Indels 943; Gaps 138;

Qy 79 SDRPQERTSYEPHPGSPVDHDSLESKRPLRQVSD--SHFORVSAAVLPVHPLEPG 136  
Db 22 SDRPSALAME-----TNPSPASPSL--LRKESEDRGDILNLRSSGA-----D 65

Qy 137 LRASADAK-----KDPAFGGKHEAPSSPTSGQPCGDQNASP-----SKLSKEELIQSM 185  
Db 66 SRASNDSSASRLPTALVGLSEAEKQHIMSVMSNRNTSPWTSRRCSSALOMLPEVDNL 125

Qy 186 DRVDRE-----IAKVEQO-----ILKLLKKQO-----QLEEEAAKPEPEKP 222  
Db 126 SEAEKEHQTILERAESKTPFMKVPKMKQISSRTESTNSRVSEGDIEEVENEVQRKKT 185

Qy 223 VSPPPVEQKHSIVQIYIDENRKAEEAHKIFELGPKVEL----PLYNQPSDTKVYHEN 278  
Db 186 IEEPIVEIPSAVT-----PRNLRV-----IPPIAISHPTPPHSAKTDTSRHS 232

Qy 279 -----IKTNQVMRKLILFKRNHARKQREQKICORYDOLMEAWKKVDRIENNP 329  
Db 233 GSSAHSQFGSTPSISGFK--IPFDK---AKTATETLVKEIKDEVIVEVDK--DKTEKP 285

Qy 330 RRKAKESKTREYKQPEIRKQREOQORFORVQORGAGLSATI-----ARSEHISEIID 385  
Db 286 EPNVSNELTAE-----EFEHRRINEM-----AGIDEIQQPPIAQRERKSSVWS 330

Qy 386 GL-----SEQENNEKO--NRQLSVTPPMFADAEQRVKFINMGLMEDPMKVYKD 433  
Db 331 GLKNMFGVKHDESELTTEEKEHRIOWS-----LMAEKM----- 364

Qy 434 RQFNWVTDEKEIKFKQFIQHPKN--FGLIASYLERKSVPCVLYLYLTKKNENYKALVR 492  
Db 365 -----DEELDEQ--SKPKSTFGL-----KNF----- 384

Qy 493 RNYGK-----RRGRNQIARPSQEEKVEEKEEDKAETKEKKEEED 539  
Db 385 --FCKATQSVNHARDSVYKVNQO--SKQSLSGLTQELDNIAQATSAQOESQELTQD 440

Qy 540 EKEDSKENTKEKKIDGTAETEETEREQATPRGRKTANSQGRKRITRSMTNEAAAAA 599  
Db 441 EIDHIA-----RINAAAEDEFQAVN-----MNOG-----LSQAEKDHIARIE 479

Qy 600 AAAATEEPPLPPPPPEPISTEPVETSRWTEEMEVAKKGLVEHGRNWAALAKVGTKSE 659  
Db 480 AWAADDQSOKIIVPPPSIIQKD-IELSM--EEMD-----HIARTAAMAD----- 520

Qy 660 AOCKNFYFNKRRHNLNLQOHKQKTSRKPREEDVSQCE-----SVASTVSAQED 711  
Db 521 ---EDF-----THPVKGAVALIYDENYSRDDGCAIDRFPASTATPVFAQPS 562

Qy 712 EDIEASNEENPEDSEAVKPSDESPENATSRGNTPEA--VELEPTE-----TAPS 762  
Db 563 E-IELS-EEEREHISRTAAAEEDFN-----TPYVSTHTQOIEIEELTEEEKEHARIAM 616

Qy 763 TSPSLAVPSTKPAEDSEVETO---VNDSISAETAQMDVDQOEHSAEGRSCVCDPPPAK 819  
Db 617 ASEDINAPSPFIQOEQRVATMPVINYHVEPTLSOEI--HIARTAM-----ATE- 667

Qy 820 DSDVDEVVRPENHASKVEGDNTKRDLDRASEKVEPRDEDLVAAQIINAQPEQSDNDS 879  
Db 668 DNTDLQT--LPTPQVKQNEPELSQE-EIDHITRIAAAMANEDEFGMYPYVSEHPAPV--ES 722

Qy 880 SATCSADEDDVDGEPRQRMFPMDSKPSLLNPTSLVSSPLKPNPLDLPO--LOHRAAVI 937  
Db 723 ELT---QEELDHIAKITGMASMDIS-TLPPPTG-----KPSETSLTQEELDHIA 770

Qy 938 PPMVSTPCNIPIGTPVSGYALYQRIKAMHESALLESQRQEQIDLECRSSSTPCGTS 997  
Db 771 E--MASAEYDVP-----KIFPELTPQEELDHIAKITAAM-- 805

Qy 998 KSPNREVEVLQAPHLITNLPVGRVRLTTPRTPPPPLPSSKTTVASEKPSFIMGSI 1057  
Db 806 -----QDVQLPATQRSITH-----KVLSTPP--PPP----- 830

Qy 1058 SQGTPGIYLTSHNQASTQETPKPSVGSISLGLPQQQSAKSATLPY-----IKOEE 1109  
Db 831 -----SKHFODLTQEE-----LHIAERIAEMADMYNTPTAEPVQDEE 870

Qy 1110 FSP--RSONSOPEGLLVRAQHEGVVRGTAGATQEGSITRGTPTSKISVESIPSLRGSTQ 1167  
Db 871 EEPITESGDSATGADIFDEQEDASSGASGDFDNNAQ--VLTSGFSPDRVTSAPALDT 928

Qy 1168 GTPALPQTGIPTEALVKGSIISRMPIEDSSPEKGRREEAKSGHVIYEGKSHILSYDNTKN 1227  
Db 929 EEOGP-----IMAKTVPSTPSADSMASRKSS--YD-----IRSEIRQ 968

Qy 1228 AREGTRSPRTAHEIS-LKRSYVESVEGNIKQMSRESVPSAPLEGL-----ICRALPGS 1281  
Db 969 EESDIGNKWEQLEQSFMRQSIHDEEDV--GHEINTDVEEPFLEYVEDQLHLEGIDVES 1026

Qy 1282 PHSDLKERTVLSG---SIMQG-----TPRATTESFEDGLKYP 1315  
Db 1027 QHHEFTTSSAFFGTGRSIGEGEKRYGDDAVEQOKLENYEEEEKTKSSREAFDDGFETQ 1086

Qy 1316 KQ--IKRESPPIRAFEAGAITKGP-----YDGIITIKEMGRSTH-EIPRODILTOE 1363  
Db 1087 REESLRAQOQTPIDSLPGSRMLKRPNGFLSNIAINDAINKAKEAGSQIAAVPIKPSSSS 1146

Qy 1364 SRKTPVVQSTRPIEGSISOGT--PIK-----FD 1391  
Db 1147 NIVNNVFFSSK-----SSTSLGTSAPTAKSIPQIGIPMDGLSEERQKIMSVMAAADFD 1202

Qy 1392 NNSGSAIKHNKVLITGPSKLSRGMPPLEIVPENIKVVERGKVEDVKA-----GETVR 1445  
Db 1203 DS-----VNNVKFSTSGSSNIPAGM-----EDLSEAEKIMSVMANEMEGARFP 1249

Qy 1446 SRHTSVSSGSPVLURSTLHEAPKAQLSPGIYDDTSARTPV-----SYQNTMSRG 1495  
Db 1250 PPSQIPIRSPSVMSSSI---MSELPPGLDLSDDERMKIMAVMAEADMQNVKPIARG 1305

Qy 1496 SPMNRTSDVTIPN-----KSTNHERKSTLTPTQRE 1527  
Db 1306 PPPMPPSTSM--IPPMGELSEERQKIMSVMANAETDSSSVITSRQPSRSPSVARMQ 1364

Qy 1528 SIPAKSPVPGVDP-----VSHSPDPHRRGSTAGEVYVWSHLPTQLDPAM 1572  
Db 1365 LMPQQAIPPIPPGLEGLSDEERHTINSVMAEAEFEES--SQVPSRQSPSPSFVNQQ 1422  
Qy 1573 PFHRALDPAAYLFFQRLSTPGY-----PSQVLYAMENTROTILNDYITSOQMVN 1627  
Db 1423 SFH-----PIPSFEPIVPPGLEDL--SEERQKIMS-----VMNA 1456  
Qy 1628 RPDVARGLSPREQPLGLPYPATRGLIDLNTNPPITLPHPGGTSTPPMDR---ITYIPGT 1684  
Db 1457 EVEESRQLPSQSPRSPSVA-----MIQAPVPIIP-SCMEDLPEAROKIMSVMAEA 1509  
Qy 1685 QIT--FPPRPYNASMSPGH--PTHAAAASAERER--ERERERERERIAAASDLYL 1738  
Db 1510 EIQNFVPSRSPSNYSQMPVPVIPHGLEDLSEARQKILSVMAEAFIDSAKIPSRSTSY 1569  
Qy 1739 R-----PGSEQP-----GRPGSHGVYVPSVPSVTOETMLQORPSVFGT 1778  
Db 1570 MPPPLQMSQPEITITGLEHSSSEADMEFGDRSSRSHQVIPPGLLENSEERQOIMSVMAHA 1629  
Qy 1779 NGTSVITLDPDTAOL-RIMPLPAGGSPISQGLPASRYNTAADALAALVDAASAPQMDVS 1837  
Db 1630 ELESII-----PSGHIDQISLPRGHTGFK---PAGIVN--EDEL-----FETERKQREES 1675  
Qy 1838 KTKESKH-----BAARLEENLRSR----- 1856  
Db 1676 PTRESGYATSYERELAMGDEERMDGLEDIIRIREGARSRRDRDVLHRRREEDPEVH 1735  
Qy 1857 -----SAVSEQQLEQKTLVEKRSV--OCLYTSAFP----- 1888  
Db 1736 TPESSTAVTVDVPSVPTENVPEKQDFDFTYSDSRFAEIVQMDEEAGSLQOKV 1795  
Qy 1889 SGKPO-----PHSSVYVSEAGKDKGPP--PKSKYE-----EELRTGK 1924  
Db 1796 DEKPRMWTVDGDESELPHQDFVNEPTTKTSDFPKPTDEVFKPSEIQIRIVTKN 1855  
Qy 1925 TTITAAINFIDVITRIQTASDKDARERSQSDSSSSLSH-----RYE 1967  
Db 1856 HDVDMDEIDYDVIATEAPSSVSQRR--QPVDSETSVKSRTIQRGTPKPPMIKITVEE 1912  
Qy 1968 TPSDAIE-----VISPASSAPPQEKLOTYQVPEVVKANQANENDPTQYEGPL 2014  
Db 1913 TKSDSESCSEDEEYDPRVVAAPTAPTFE-----EVENERIKQ----- 1954  
Qy 2015 HHYRPQSPSPQOOLPSSQAEGMGQVPRHRLITLADHICQIITODFARNQVSSQTPQ 2074  
Db 1955 -----EELGKVLQOI-----MAFG-----EVANDEFDQWAKTITSQ 1987  
Qy 2075 QPPTSTFQNSFALYSTVPTVTKTNSRYSPEQAQSVHHQRGSRVSPENLVDKS----RGS 2131  
Db 1988 TPSTSTKPTVTPAKRSDPI-----PIAPSQRSEIEBE-----RIRTEALEEEEFYRHG 2037  
Qy 2132 RFG-KSPERSHVSSEPEYPISPQVVPVHKEQDSLILLLSQGAEP-----EQNRDARS 2184  
Db 2038 NPFLESPEDEVSIN-MEDVYEAELARYESANO-----TMRPQGVYITIDESEDDGTL 2092  
Qy 2185 PGSISYLPFFTKL--ENTSPMWKSKQIEFR-KLNSGGGSDSMAAAQPGTEIFNIPAV 2241  
Db 2093 SNSESRLVAREKRLMDKTAADSLMAKYQKMKVQAKQTITASSVTSVTPATAYAINFSDL 2152  
Qy 2242 TTSQSVSRGHSFADPASNGLIEDI-----IRKALMGSPDKVEDHGVMVMSQ----- 2288  
Db 2153 KTS-TRTTDSNSYFETTKNIPALEIKDPKKDIPPEISASIDKTMAEVDALLGQVYTNKA 2211  
Qy 2289 -----PMGVVPGTANTSVTSGETREEDGDPSPHSGGVCKPKLISKNSRKS 2337  
Db 2212 IPNLICFQNSFNHPHSSASTATADDL-----ILLKNNSSSPSF 2253  
Qy 2338 --PIPGQGLYTERPSSVSSVSGDYH-----RQTPGWAWEDRPSSTGSGTFPY 2385  
Db 2254 LLPLQ-SSVLG-----SOLDSVYRDDNERNENETSPRGLKRSFGML---LPSPSTSTSIFF- 2304  
Qy 2386 NPLWMLSSPTPTPIACAPSAVNOAAPHQONRIWREPAFL-----LSAQYETLSD 2437

Db 2305 -----MPPTAAESVGAAGATTATSMFEGGISVADPPPTMDGLTNSYKWLQN 2349  
RESULT 15  
T30826  
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse  
N;Alternate names: alpha-NAC protein  
C;Species: Mus musculus (house mouse)  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 24-Nov-1999  
C;Accession: T30826  
R;Yotov, W.V.; St-Arnaud, R.  
Genes Dev. 10, 1763-1772, 1996  
A;Title: Differential splicing-in of a proline-rich exon converts alphanAC into a mus  
A;Reference number: Z20889; MUID:96312450  
A;Accession: T30826  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2187 <YOT>  
A;Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1  
C;Genetics:  
A;Gene: Naca  
A;Map position: 10  
A;Introns: 24/1; 1996/1; 2050/3; 2050/3; 2142/3; 2183/3  
A;Note: differential splicing converts alphanAC into a tissue-specific DNA-binding ac  
C;Keywords: alternative splicing; DNA binding; transcription factor  
Query Match 3.2%; Score 408; DB 2; Length 2187;  
Best Local Similarity 20.5%; Pred. No. 2.6e-08;  
Matches 404; Conservative 261; Mismatches 780; Indels 524; Gaps 92;  
Qy 504 QIARPSOEKVEEKEDEKAEKTEKKEEKEDEKDEKDEKENT-----KEKDKIDGTA 558  
Db 497 QAGLPTKRTDTLQPLALIKESPPSSQSSASLEVLSEDTVTYKTTGGPVPVRPAIGVA 556  
Qy 559 EETEEREQATPRGRKTAN--SQRRKGRITRSMTNEAAAAAATAEPPPLPPPPPE 616  
Db 557 TTTSLRADSPPAVIRADSCVSPNTVSQPLKRSVTDPAAPRTAKNTA-----PSTTSPLV 611  
Qy 617 PISTE--PVETSRTEEMEVAKKGLVEHGRNWAIAKMGVTKSEACKNFYF----- 667  
Db 612 PLASEGCPVASS-----MALSPQN-----ASVSETALALSPEITPKSVFPDPPLAE 657  
Qy 668 -NYKRRINLNLQHQKQTSRKPREERDVSQ-----C-----ESVASTVSAQEDIEDASN 718  
Db 658 ISFSNARKVDVAVSHWESSGSSRQGHDPDASVYAKGTVVCLADSLDTSVSASKGALSAS 717  
Qy 719 EENPEDESE-----VEAVKPSD--SPENATSRGNTPEPAVELEPTTETAP-----ST 763  
Db 718 SPLYPLEVSFLPEAGLAVQGPGLSLNKLSPTPPSKSGAPVSTGAPSPKGPATVPTESS 777  
Qy 764 SPSLAVPS-TPAEDESVEITQVNDISIAETAEOMD--VDQOEHSAAEGSVCDPP----- 814  
Db 778 ISSQVFAEILPSPQKTEVTASRLISAVOSPKVDPTMSQVPTPSPKRTSATAVPKDTSA 837  
Qy 815 -----PATKADSDVDEVVRVPRPENHASKVEGDNTKERDLDRASEKVEPRDELDYVAQOI- 866  
Db 838 TSLSKSVPAVYTSLSPPKAPVAPSNEATVPTPE--IPTSLKNALAAATPKE---TATSIP 892  
Qy 867 NAOPEPOSNDSSATCSADSDVDGEPERQRMFMDKPSLLNPTGSLVSSPKLPN--P 924  
Db 893 KVTSPSPQKTPKSVS-----LKGAP-----AMTSKKAT-----ETAASKDYSPSQFP 934  
Qy 925 LDLPLQLOHRAAVIPPMVSCYPCNIPIGTPVSGYALYORHIKAMHESALLEEORQROQID 984  
Db 935 KEVPLLOH-----VPP-----TSPPKSPVSDTLSG-ALTSPPPKG---PPATLAB----- 975  
Qy 985 LECRSSTSPCGTSKSPKREWEVLQAPAPHLITNLPGLVRLPTTRTPPPPLIPSSKTTV 1044  
Db 976 -----TPIYPKSKPAAKSKTTPA-----TPSPEGV-----TAVPLEIPCSKKAP 1016  
Qy 1045 ASEKPSFMGSGISQGTGTYLTSHNQASVYQETPKPSVSGSISL--GLPROQESAKSATL 1102



Db 1017 KTAAPKSSATSSSKRAPKT-----AVSKEIPSKGTAVPLEISLP-LKETSKSAT- 1066  
QY 1103 PYIKOEESPRSONSQPELLVRAQHEGVVRCGTAGAIQ--EGSITRGTPTKISVESIP- 1159  
Db 1067 -----PGEKASAPKSKPK-----TAGPKETPPGGVTAVPPEISLPPKETPQ 1108  
QY 1160 -----SLRGSIQGTAPALPOTGIPTEALVKGSISSRMPIE-DSSPEKREAAASKGHVY 1212  
Db 1109 NATPNESLAASQKRS-----PKTSVPKET-PPGGVTAMPLEIPSAPQKAPKATVPKQIPTP 1164  
QY 1213 EKGSHIILSYDNIKNARCGTRPTAHEISLKRVSVEGNIKQGMRSRESPVSAPLBGL 1272  
Db 1165 E-----DAVTILAGSLSPKASKTA-----APKEAP-ATPPSGV 1198  
QY 1273 ICRALPRGSPHDLKERVLSQIMQGTTPRATTESFEDGLKVPKQIKRESPIRAFEGAI 1332  
Db 1199 I-----AVSGEI-SPSKTKST-----AAPKENSATLPPKRSPTAA 1235  
QY 1333 TKGPK-----YDGITTIKMGRSIHEIPRODILTQESRKTPEVVQSTRPIIEGSIQGTPIK 1389  
Db 1236 PKETATPSSEGTAV-----PSEISPSPT-----PASKGVPT 1269  
QY 1390 FDNNSGQSAIKHNKSLITGSKLSRGMPLEI-----VPENIKVVVERKYEDV-----KA 1440  
Db 1270 LTPKGAPNALAESPAS-----PKKVPKTAAPETSTTPSPQKIPKVAGPKEASATPPSKKT 1325  
QY 1441 GETVRSRHTSVSSG-----PSVLRSTLHEAPKAQLSPGIYDDTSARRTPVSVQNTMSRGS 1496  
Db 1326 PKTAVPKTSPSEGTAVPLEIPSPKAPKT-----AAPKETPAP-----S 1368  
QY 1497 PMMNTSDVTIPP-----KSTNHERKSTLTPTQRESIPAKSPVPGVDVPSVSHSPDPH 1550  
Db 1369 PEGATTAPVQIPPPSRKSGKAGSKETPTTSPGCVTAAPLEIPSSKTKSMASPRE-- 1426  
QY 1551 HRGSTAGEVYVSHLPTQLDPAMPFHRALDPAAAYLFQRLSQSPGPGYPSQYQLYAMENR 1610  
Db 1427 -----TLVTPS-----SKLSQTVG-----PK 1443  
QY 1611 QTILNDYITSOQMVNLRPD--VARGLSPREQPLGLPYPATRGIDLTNNPPTILYPHP 1667  
Db 1444 ETSLEG-----ATAVPLEIPPSHKKAPKTVDPKQVPL-TPSPKD-----APTLEASP 1490  
QY 1668 GG-----TSTPPMDRITYIPGTQITPPRPYNSASMSGHPTHLAAASAERER----- 1716  
Db 1491 SSPKAPKTAAPPSESVTVTPPERPATPKAGSTASKVPVPAETQEVAVSSRETPTVPA 1550  
QY 1717 -----EREKERERERIAAASDLYLRG-SEQGRPGSHGVRS-----PSPSVRTQ 1764  
Db 1551 VPPVKNPSSHKTKSTIKLEKAPATLPPSPKIPSSKKAPRTSAPKEPPASPSIKPV 1610  
QY 1765 ETML-QORPSVFOGTNGTSVITPLDPTAQLRIMPLPAGGSPISOGLPASRYNTAADALA- 1822  
Db 1611 TSLAQTAPPSLQKAPSTI-----PKNL-----AAPAV-----LPVSSKSPAARAS 1655  
QY 1823 ALVDAASAPOQ--DVSKTESKHEAARLEENLRSRAAVSEQOOLEKTLVEKRSYQC 1880  
Db 1656 ASLSPATAAPQATPAKEATTIPSCKKAATETPIETSTA-----PSLEGAPKETSETSVSK 1710  
QY 1881 LYTSAPPSGAPQPHSSVYSEACKDGPPKSYEEELTRGKTITTAANFIDVITRQ 1940  
Db 1711 VLMS-----PPKASSSKRASTLPLATTLPSLKEASVLSPSTA 1747  
QY 1941 IASDKDARERSSQSDSSLSHRYETPDSAEIVISPASSPAPPOKLOIYQEVVKAN 2000  
Db 1748 TSSGKDSHI--SPVSDACSTG-----TTPQASEKLPSSKKGPTAFTLEMAAPAPESALAI 1800  
QY 2001 QAEINDPTRYEGPLHHYRPOQESPSPQOQLPPSSQAQMGQVPRTHRLITLADHICOLIT 2060  
Db 1801 TA-----PIQKSGANSNSASSPKCFDPSKDKTKGLPSAVALAPQTPVPEKDTSKAETILL 1857  
QY 2061 QDFARNOVSGTPOOPTSTFTQNSPNSALVSTPVTKTSNRYSPSQASQVHHQRPGRSVS 2120

Db 1858 VSPAKGSDCLHSPKGPVGSQ-----VATPLAAFTSDKVPPEAVSASV----- 1899  
QY 2121 PENLVDKSRGSRPGKSPERS-HVSSEPYEPISPPQVPPVHKEQDLSLLLSQSGAPAEQR 2179  
Db 1900 -----APKAPAASLTAPSPVAPL-PPKOPLESAPGSVLESPSKLPVPABED 1947  
QY 2180 N-----DARSPGS-----ISYLPSEFTTKLENTSPWYKSKQEIFRKLNSGGG---DS 2224  
Db 1948 ELPLPPIPEAVSGGEPFQPILVNMPA--PKPAGTAPAPASAKQPVLK--NNKSGSTESDS 2003  
QY 2225 DMA-----AAQPGTEIFNLPAVT-----TSGSVSSRGHSIFADPA-SNLGLBEDI--- 2266  
Db 2004 DESVPELEQDSTOTATQOAOAAAAEIDEPVSKAKOSREKKARKAMSKLGLRQVTGV 2063  
QY 2267 ---IRKALMSFDDKVEDHGVMSOPMGVPGTANTSVTSGETRREE 2311  
Db 2064 TRVTIRKSNILF-----VITKP-DVYKSPASDTYIVFGEAKIED 2102

## RESULT 16

T20532  
hypothetical protein F07A11.6b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20532; T27777  
R:Palmer, S.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: T219287  
A:Accession: T20532  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2722 <WIL>  
A:Cross-references: EMBL:Z66511; PIDN:CAB54211.1; GSPDB:GN00020; CESP:F07A11.6b  
A:Experimental source: clone F07A11  
R:Gajadaty, S.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: Z20417  
A:Accession: T27777  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2722 <WI2>  
A:Cross-references: EMBL:Z69904; PIDN:CAB54502.1; GSPDB:GN00020; CESP:F07A11.6b  
A:Experimental source: clone ZK20  
C:Genetics:  
A:Gene: CESP:F07A11.6b  
A:Map position: 2  
A:Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3

Query Match 3.2% Score 405; DB 2; Length 2722;  
Best Local Similarity 18.1% Pred. No. 4.4e-08;  
Matches 442; Conservative 340; Mismatches 823; Indels 836; Gaps 111;

QY 3 SSGYPNQGAFSTEQSRYPHSHVQYTPNTRHQEFAVPDYRSSHLEVSQASQLLQOQOQ 62  
Db 687 TSGRP-----SLDESR-----TNRLSPDSTHHPAELA---QRSHSLCIGPMTPTSPPTS 733  
QY 63 QQLRRRSLLESEFH-PGSDRPQER-----RTSYEPHPGSPVD-HDSLESK-RPRLEQ 113  
Db 734 Q-----PLLNVNTHLPQTSPQSTSGGITTPKSSQPP--PLMSPVSRHNSMSSTGRP----- 782  
QY 114 VSDSHFORVSAVLPLVHPLPEGLRASADAKKDPAGFGKHEAPSPISGQCGDDONA-- 171  
Db 783 -----ASITLHR-----QSVFFPDVSIPIPPPI--PPTHENMAWR 817  
QY 172 --SPSKLSKEELIQSDVRDREIAKVEQIILKKKQOOLEEAAKPEPEKPSPPPPVE 229  
Db 818 GTPPSRSSETMVLPR--SPFGCTPIQNLLTM-----PIVPPPP-- 853  
QY 230 QKHSRIVQIYDENRKAEEAHKIFEGLPKVEL---PLYNQPSDTKVYHENIKTNQMR 286  
Db 854 ----HLIAATSTGTHSVSSSAHST-----PRHSISGTPVHCEPS-----NSKTSQ--- 894





|    |      |                                                                 |      |
|----|------|-----------------------------------------------------------------|------|
| Db | 2016 | DSAVDLSLKH5YSLGFADGRYLGGQGLQYGSFTDRLRHPTDLLSHPLPLRRYSVSSVNSYSD  | 2075 |
| Qy | 1505 | VTIPPNKSTNHERKSTL---TPTQRESIPAKSPVPGVDPPVYVSHSPFDPHHRGSTAGEVY   | 1560 |
| Db | 2076 | HRYPGRGDVAFQOEASLAQYSATTAREISRMCAALNSMD-----QYGGRGSGSGGPD       | 2129 |
| Qy | 1561 | WSHLPTQLDPAMPFHRALDPAAAAYL---FQRLSPTPGYPSQYOLYAMENR-OTILND      | 1616 |
| Db | 2130 | VOYQP-QHGPGLSAPQGLAPLRSLGLGNPTYPEGOPSPGNLAQYGAASQATAVRQLLPS     | 2188 |
| Qy | 1617 | YIYSQ---QMQVNLBPDVARGLSPREQPLGLPYPATRG-----IIDLTN               | 1657 |
| Db | 2189 | TATVRAADGMIYSTINTPIAATLPIITQSPASVLRPMVRGMYRPYVSGGVTAVALTSLTR    | 2248 |
| Qy | 1658 | MP-----PTILVPHPG-----GTSPPPMDRITYIPGTOIT-----PFPYPYNSAS         | 1697 |
| Db | 2249 | VPMIAPRVPLGPAGLYRYAPAPRPTIASSVPPAEGPVYLCKPATKASGAGGPPRPELPAG    | 2308 |
| Qy | 1698 | MS-----PCH-PTHIAAASA-----                                       | 1712 |
| Db | 2309 | VAREEPSTTAPAVIKEAPVAPAGPAPAPPFGQKPAGEAAGSGGVLSPASEKEEAS         | 2368 |
| Qy | 1713 | -----PRER-----PRER---EKERERERIAAASDLYLR                         | 1739 |
| Db | 2369 | QEDRQRKQBLQLEREREVELEKLRQLRQLEBERERVELQRHREEBQLLVQRELOELQ       | 2428 |
| Qy | 1740 | P-----GSEQPCRGSHGVYSPSPSVTQETMLQORPSVFOGTINGTSVITPLDPTAQL       | 1793 |
| Db | 2429 | TIQHVLUQOOOEERQAFALQREBLAQRLQLEQLQOQOQLQ-----LQLEEKQR           | 2480 |
| Qy | 1794 | RIMPLPAGGPSISQGLPAGRYNTAADAL-----AALVDAAS---APQMDVSKTKES        | 1842 |
| Db | 2481 | QKAPFPATCEAPSRGPPPAATELAQNGQWYPLTHAATFVAVAGTSGQOPREP- ---       | 2534 |
| Qy | 1843 | KHEAARLEENLRSAASVSEQ---QOLEORTLEVKRSVOCLYTSSAFSPGKQPQPHS ---    | 1896 |
| Db | 2535 | -----LHRGLPSSASDMSIQTEQWEAGRSQIKKRRHSPRLRLDACEPESGPDSTVRR       | 2588 |
| Qy | 1897 | ---SVVYSEACKDKPPPKSRYBEELTRCKTTITAAFNFDVLIITQIASDKD-----A       | 1947 |
| Db | 2589 | ADSVQTDDEGEQ-----RY---LVTRRRRTTASA---DCSVQTDDEDNADWEQPVRR       | 2637 |
| Qy | 1948 | RERGSQSSDSSS-----SLSHRHYETPSDA-----IEVISP- 1979                 | 1979 |
| Db | 2638 | RSRLSRHSDSGSKDHATASSSTTAAATARAMSSVGIQTISDCSVQTEPQLPRVSPAI       | 2697 |
| Qy | 1980 | --SSPAPPQBLQTYOPEVYVVKANQEN-----DPTROVEG-----PLHRYRP- ---       | 2019 |
| Db | 2698 | HITAAATDPKVEIVRYISAPEKTGRGESLACQTEPDGQAQVAGPQLIGITAI5PVLPGIQ    | 2757 |
| Qy | 2020 | -----QOESPSP-----QQOLPPSSQAEHGM-QVPRTHRLI-----TLADHTCQI         | 2058 |
| Db | 2758 | IVTPGALGRFEKKKPDPLEIGYQAHLPESLSQLVSRPPKSPQVLYSPVSPLSPH--RL      | 2815 |
| Qy | 2059 | ITQDFARNQ-----VSSQTPQOP-----PTSEFNQSPSALYSTVPRKTSNRY            | 2101 |
| Db | 2816 | LDTSPASSERLNKAHVSPQKQFIADSTLRQOTLPRPMKYLQRS-----LSDPKPLSPTAEE   | 2871 |
| Qy | 2102 | SPESQAQSVHHQRP-GSRVYS---PENLVDKSRGSRPGKSPERSH-----VSSEPYEP- --- | 2149 |
| Db | 2872 | SAKERFSYQHOHGGLSQSVLPNGLVRKVKRTLPSPPPPEAHLPLAGQVP5QLYAASL       | 2931 |
| Qy | 2150 | -----ISPOQVPVH-----EKQDSLILLLSQRGAEPAEQNDARESGSISY              | 2190 |
| Db | 2932 | LQRLAGPTTVPATKASLLRELDRLRVHEHSTKLRKKQALDEEKEIDAK-----LKY        | 2987 |
| Qy | 2191 | LPSFFTKELENTSPMWKSKQEIFRKLNSGG-----GSDM-----AAQPGTEIFENLP       | 2239 |
| Db | 2988 | LELGITQRKES--LAKDRGGRDYPPLRGLGEHRDYL5DSELNQLRQCCTTACQYVDYP      | 3045 |
| Qy | 2240 | AVTTSVGSVSRGHFSFADPASNIGLEDITRKALMG5FDKVDHGVVMSQPMVYVPGTANT     | 2299 |

```

Db 3046 A-----SAAVPAATPSG-----PTAFQQR-PPAPQY 3072
Qy 2300 SVVTSGTREEDGPSHSGGVCKPKLIKSNRKSIPGGYGILGTERPSSVVSHSE 2359
 : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 3073 TAGSSGPT--QNGFPAHQ-----PTYTGSTYPATYP-PGTGYPAEPLSPQAFPHPT 3124
Qy 2360 GDYHRQTPTGNWEDRRPSTGSTGTPPY-----NPLTMRLMSSTPPT--PIACAPS 2406
 : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 3125 GHYAAPT-----MPTQSADFPVOADSRAAHQKPROTLADLEQKVPTNYEVIGSPA 3177
Qy 2407 AVNQAAAPHOQNRIWEREPALLSAQYE 2433
 : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 3178 VTMSAPPETG-----YSGPAVGSYE 3199

RESULT 18
A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: A48666
R:Schluter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge, u.
J. Cell Biol. 123, 513-522, 1993
A>Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u.
ins.
A:Reference number: A48666; MUID:94043435
A:Accession: A48666
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCH>
A:Cross-references: EMBL:X65550; NID:g415818; PIDN:CAA46519.1; PID:g415819
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KH>

Query Match 3.2%; Score 404; DB 2; Length 3256;
Best Local Similarity 18.1%; Pred. No. 6.1e-08;
Matches 516; Conservative 368; Mismatches 1083; Indels 882; Gaps 128;

Qy 46 SHLEVSQASQLQQOOQQOLRR-----RPSLLSEHPGSDRPQERTSTPEPHGP 96
 : :::: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 482 SSVDDINFGDSINEGIPLKKRRRVSPFGHLRPFELDENLPN-----TPLKRGE 531
Qy 97 SPVDHDSL-----ESKRPRLEQVSDSHFQ-RYSAAVLPLVHPLPEGLRASA- 141
 : ||| ||||| : : : | : | : | : | : | : | : | : | : | : | : | : |
Db 532 APTRKSLVHHPTPVLVKIIKEQPQPSKGESIEHVVEKAQSLVISPPSPSRPTPVA 591
Qy 142 -----DAKKDPAGPGKHEAPSSPISQPCGD---DQNASPSKLKSEELIQSMRDVRE-- 191
 : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 592 SDQRRSCKTAPASSRSKOSTEVFKRGGERVATCLKRVSLR-SQHDLQMICKRRSGA 650
Qy 192 -----TAKVEQIILKUKKOQ-----LEEAAKPPEPEKPY----- 223
 : ||| : ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : |
Db 651 SEANLIVAKSWADVVLGAKQTQTKVIKHGQRSMNKRQRRPATPKKPVGEVHSQSFSTGH 710
Qy 224 --SPVPV-----EQKH-----RSIVOILIYDNRKKAEEAHKIEGLCPKVELPLYNQ 268
 : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 711 ANSPCTIIIGKAHTKRVHVPARYVLNNFISNMOKMFRED----LSGIAEMFKTPVKQE 766
Qy 269 PSOTKFVYHENIKTNQVMNRKKLI-----LFFKRRNHARQREQKIC 308
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 767 POLTSTCHIALISENLGKFOGTDGEBPLPTSESFGNVFFSQN-AAQPSDK-C 824
Qy 309 QRYDQLMEAEKKVDRIENNPRAK-----ESTREYYEKOFPEIKRQEQERFORVGQ 364
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 825 SASPLRQCIRENGNNAKTPTNTYKMTSLETKTS-----TETEPSKTVSTVNR 874
Qy 365 RGAGLSATIARSEHETSEIIDGLSEQUENNE-----KMQLSVIPPMPFDASORRVKFI 418
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 875 SG-----RSTEFRTOKLP--VESKSEETNTFIVECILRKQOKATLL-----QQR----- 918
Qy 419 NMINGLM---EDPMKVYKDRQFMVNWDHKEIFEKDQFIQHPKNFL-----IASYLKERKS471

```

Db 919 --EGMKIEERPFETYKE--NIELKENDE--KMKAMKRSRTWQCKCAPMSDLTDLAKSL 970  
Qy 472 PDCVLYLYLTKNENYKALVRNKGKRGRRNOQIARPSQEKVEEKEDKAEK-----524  
Db 971 PDELM-----KDFARGONLQTOHAKAPKSEKGIKMKPCQSLQ 1011  
Qy 525 -----TEKKEEKDEBEKED-----SKENTREKDKIDGTAET---E 562  
Db 1012 PEPINTPHTKQOLKASLGKVGKRELLAVGKFTSGETHTHREPACDGSIRTFKES 1071  
Qy 563 EREQATPRGRKANSQGR-----KGRITRSMNEAAAAA 601  
Db 1072 PKQILDPAARTVMKKWPRTPKEEAQSLDLAGKELFQTPGSESMDEKTTKIACKS 1131  
Qy 602 AATEPPPPPLPPPEPISTEPVETSRW-----TEEMEVAKKGLVEHGRNWAATAK 652  
Db 1132 -----PPPESDVT--PTSTKQWPKRSLRKADVEEFALRKLTSPSAGK--AMITP 1177  
Qy 653 MVGTKSEAQNFFNYKRRHNLNLOQHOKTSSRKPREE-----RDVSQCESVASTVSA 708  
Db 1178 KPAGGDEKIDKAFMGTVPQKLDLACTLPGSKROL--QTPKEKAQALDLAGKELFQTPG- 1235  
Qy 709 QEDIEDIASNEE-----ENPDESEVAVKPSEDSPENATSRGNETPAVELEPTTETAPST 763  
Db 1236 --HTEELVAAGTKTIPCDSPQSDPVDPTSTKQRPKRIRKADVEG--ELLACRNLMPSA 1292  
Qy 764 SPSLAVPTKPAEDESIVETQVNSI-----SAETAEQM-----DV 798  
Db 1293 GKAMHTPKSPGEEKDIIIFVGTVPQKLDLTENLTGSKRRPOTPKKEAQALEDLTGFKEL 1352  
Qy 799 DQEHSAEGSV-----CDPPATKADSDVVEVRVPE-----830  
Db 1353 FQTPGHTEEAAGKTKMPCESSPPESADTPTSTRQPKTLEKROVQKELSALKLITQ 1412  
Qy 831 -----NHASKVEGONTKERDLDR--ASKVEPRDEDLVVAQOINQAORPEPDNDSSATC 883  
Db 1413 TSGETHTDKVPGGEDKSINAFRETAOKKLDP-----AASVTGSKRHPKKEKAQPL- 1464  
Qy 884 SADEVDGCEPERORFMPSKLLNPTGSLVSSPLKPNLDLPOLQH-----932  
Db 1465 ---EDLAGKELFQTPVCTDPTTHEKTKIACKS--OPDPVDTPTSSKPOSKRSLRKVD 1519  
Qy 933 -----RAAVIPPMVSCTPCNP--IGTPVSGYALQ-----RHIAKMH 968  
Db 1520 VEEEPFALKRTPSAGKAMHTPKPAVSEKNIYAFMGTPVQKLDITENLTGSKRLQTPK 1579  
Qy 969 ESAL-----LEQORQOEQIDLECRSS-----990  
Db 1580 EKAQALDLAGKELFQTPRGHTEESMTNDKTAACKVACKSSQPDLDKNPASSKRLKTSIGK 1639  
Qy 991 -----TSPCGTSKSPNREVEVLQAPHOLI---TNLPEGVR 1023  
Db 1640 VQVKEELLAVGKLTQTSGETHTHTEPTGDKS-----MKAFWESPQKQLDSAAALTSKGR 1695  
Qy 1024 LPTTRTPRPPPLIPSSKTTTVASEKSPIMGSGISQSGTPGTILTSNOASYTQE--TPKPS 1082  
Db 1696 QLRT-----PKGSEVPEDLAGFI---ELFQ--TP-----SHTKESMTNEKTTKVS 1736  
Qy 1083 -----VGSISLGLPQOQSASATLPYIKQEFSPRSQNSQPEGLLVAQRAHGVVVG 1134  
Db 1737 YRASQPDLDVDTPTSSKPOPKRSLRAD-----TEBEFLAFRQTPSAG---KAMH--TKP 1787  
Qy 1135 TAGAIOESITRGTPTKSIVESIPSLRGSITQGTALPQFG--IPTALVGSISRMPLE 1193  
Db 1788 AVGEKDLNTELGTPVQK-----LDQPGNLP-----GSNRL-----1819  
Qy 1194 DSSPEKGRGEEAASGHVIECKSGHILSYDNKNAEGTRSPRTAHETSLKRSVESBGN 1253  
Db 1820 -----QTRKEAQ-----ALEELTGFRFLFQTPCNDNPTADEKTKKI---1857  
Qy 1254 IKQGMMSRSPVAPLEGLICRALPRGS---PHSDLKERTVLSGSIQGTPTATSPED 1310

Db 1858 -----LCKSPQSDPAD-----TPTNTKQRPKRSLSLKADVEEFL---AFKRLTPSAGK 1902  
Qy 1311 GLKYPKQIKRSPPIRAFEAGIT-----KGKPYDGTITTKEMGR 1349  
Db 1903 AMHTPKAAVGEKDLNTEFGTPVEKLDLGNLPGSKRRPOTPKKAKALEDLAGKELFQ 1962  
Qy 1350 S---THEIPRODILTOESRKTPVQVSTRIEIGSISQCTPIKFDNNSGQ-----SAIKHN 1402  
Db 1963 TPGHTEESMTDDKITEVCKSP-----QPDVPKTTSSKQRLKISLGKVG 2007  
Qy 1403 VKSLITGSPKLSRGMPLLEIVPENIKVVVERGYEDVAGETVRSRHTSVVSSGSPVLSRT 1462  
Db 2008 VKEEVLVPVKLITQ-----TSGKITQT--HRETAGDKSI--KA 2041  
Qy 1463 LHEAPKAOLSPCIYDDTSAR--RTPVSQNTMS--RGSFMMNRTSDVTIPPNKSNHERRK 1518  
Db 2042 FKESAKQMLDPAHYGTGMRWPRTPKEEAQSLDLAGKELFQTPD-----HTEE 2091  
Qy 1519 STLTPQRESIPAKSPVP--GVDPVVS--HSPFDPHHRGSTAGEV-----YWSHLP 1565  
Db 2092 ST--TDDKTTKIACKSPPESSMDTPTSTRRRPKTPLGRDIVEELSALKQLTQTTHTDKVP 2150  
Qy 1566 TOLDPAMPFHR-----ALDPAAAYLFORQLSPTPGYSQ--YQIYAMENTROT--ILNDY 1617  
Db 2151 GDEKGINVRETAKOKLDPAASVTGSKRQ--PRTPKGAQPLEDLAGKELFQTPVCTDK 2209  
Qy 1618 ITSQOMQVNLDPDVARGLSPREOPLGLPY-----PATRGI-----IDLNMPTT--1662  
Db 2210 PTTHEKTKTI---ACR--SPODPVGTPTIFKPOSKRSLRKADVEEESLALRKRTPSVGK 2264  
Qy 1663 --LVPHPGCTSTPPMDRITYIPGTQITPPRPYNASPSGPHPTHLAAAAAERERERER 1720  
Db 2265 AMDTPKPAAGDEKMAFMGTVPQKLDLP-----GNLPG-----SKRWQTPK 2307  
Qy 1721 EKERERERIAAASDLYLRPQSEQPRGSGHYV--RSPSPSVRTQETMLQORPSVFOGT 1778  
Db 2308 EKAQALDLAGKELFQTPGDKTDEKTTKIACKSPQDPVDTPTASTKORPK-----2361  
Qy 1779 NGTSVITPLDPTAQLRIM--PLPAGGPSISQGLPA---SRYNATAADALAAALVDAASAP 1832  
Db 2362 ---RNLKADVEEFLALRKRTPSKAKAMDTPKPAVSEKNIINTVETVPQKLDLGNLP 2418  
Qy 1833 QMDVSKTESKHEAARLEENLRS-----RSAAVSPQOOLEQKTEVERKRSVOCLYTSSAF 1887  
Db 2419 --GSKRQOTPKKEAALDVLGFKELFQTPGHTESMTDDKITEVSKS-----2466  
Qy 1888 PSGRPOPHS--SVVYSEACKDKGPPPKSYEE-----LRTRGKTTITAAFINDIVITR 1939  
Db 2467 ---QPESFKTSRSKORLKIPLVKVDMKEEPLAVSKLTSTGTTQHTB-----PTG 2517  
Qy 1940 QIADKADAREGSSOSSSSSLSSHR-----YETPSDAIEVIS 1977  
Db 2518 DSKSIKAFKESPKQILDPAASVTGSRROLRTRKEKARALEDLVDFKELFSPAGHTEESMT 2577  
Qy 1978 -----PASSPAP-----POEKLQTYQPEVVKANQAEENDPTROYEGLP 2014  
Db 2578 IDKNKTIKCPKPPPELTDATSTKRCPKTRPRKEVKEELSAVERLTQTSQGSTHHTKEPA 2637  
Qy 2015 -----HHYRPOQESPSQOOLPSSQAQEMGQVPRTHRLITLADHTCQIIT 2060  
Db 2638 SGDEGIKVLKORAKKKNPVEEPPRRRPRAPKEKAQPLEDLAGTELSETSGHTEQESIT 2697  
Qy 2061 QDFARNQVSSOTPOQPTST-----FQNSPSALYSTPVRTKTSNR--YSPE 2104  
Db 2698 AGKATKIPCESPPLEVDTTATSKHLRTRVQKVQVKEEPSAVKFTQTSGETTADKPEPA 2757  
Qy 2105 SOAQSVHHQRGSRVSNLVDRGSRSPGKSPERSHVSSEPIEPSPPQVVPVHEKODS 2164  
Db 2758 GEDGKIKALKESAKQTPAPASVT--GSR--RRPRAPRESAQAIEDLAGFKDPAAGHTEES 2814  
Qy 2165 LLLLSQGAEPAEQNDARSFGSISYLSFTFKLENTS-----PMVSKKQEIFRKLNS 2218  
Db 2815 M-----TDDKTTKIPCKSS-----PELEDTATSSKRRRPRTRAQVKEELLA 2857



Query Match 3.1%; Score 391.5; DB 2; Length 2282;  
Best Local Similarity 18.8%; Pred. No. 1.2e-07;  
Matches 468; Conservative 294; Mismatches 886; Indels 843; Gaps 115;

Qy 2 SSGYPPNQGAFSTEQSRYPHPSVOYFTPNTRHQOEFAVDPYRSHLEVSOASQLLOQOQ 61  
Db 33 SSAPY---GSGTTAPSSATQELLATPFGSPSQE-----KTGQ 69

Qy 62 QOQLRRRLSLLSEPHGSDRQERRTS-----YEPFHGPGS----- 97  
Db 70 QOKPARRPSIEASVHI--SQLQHLPLTAPFMPGPKPEHLLGSTWQLVDPMRPGSGSVA 128

Qy 98 PVDHDSLESKRPRLEQSDSHFORVSAVLP-----LVHPLPEGLRASADAKD 146  
Db 129 PGSHPO-----SQLPSSH-----ASILPPEELPGIPKVFVPRPSQVSLKPAFAEHK- 174

Qy 147 PAFGKGHEAPSP-----ISQPCGDDQNASPSKLSKEELIOS----- 184  
Db 175 -----KERKPOKPGKICQYCSRPC-----AKPSVLQKH-----IRSHGTGRPYPCGPGGSF 222

Qy 185 -----MDRVDREIAKVEQOILKLKKKQOOL 209  
Db 223 KTKSNLYKHRSKSHAHRTKAGLASGSSSEMYPPGLEMERIPGE--EFEE---PTGESTDS 277

Qy 210 EEEAAKPEPEKPPVEQKHSIYOIYD--ENRKAKEAHHKIFEGLPKV----- 261  
Db 278 BEETGAASGPSTOVLKPK--KHPLLSLLSYSGSHGSSQERCISQSSTGSPLEDPAFP 334

Qy 262 -----ELPLYNQPSDKYHNIKNQVMRKLLILFKRRNHAKRQOKTCQRYDQIME 316  
Db 335 AEASSEHPLSHKPDTHTIKOKLALRLSERKKLI---EEOFTLSPGSKGSGTESYFSRSE 391

Qy 317 AWEKKVDRIENNRKAKESKREYIEKQPEIRKQEQOERFORVORGAGLSATIA- 375  
Db 392 SAEOQV-----SPPWTKAK-----SYAEIIF-----GCKRGIRGORTSMLASTSQP 432

Qy 376 -----SEHEIS-----BIIDGLSEQEN-NEKOMRQLSVIPPMFMDAPQRRVKFTNM 420  
Db 433 LPLSSEDKPSLVPLSPRTQVIEHITKLITINE-----AVVDTSEIDSVKPRSSILTR 486

Qy 421 NGLMEDPMKVKYKDFQFNVWTDHKEIFKDK---FIQHPKNFGLIASVLEKSPVD--CV 475  
Db 487 RSSVESP-----KSLYRDSLSHGEKTKQEQSLLSLQHPSPSTHPVPLLRSHSMPSA 542

Qy 476 L-----YVY-----LTKKNENKALVRNRYGKR--G 500  
Db 543 ISTHHHTFRGSYFDDHVDADPEVFSRNPVPTSHPRMLKRAHAIELPLGGEYSSEEPGS 602

Qy 501 RNOQIARPSQBEKVEEEDKAETEK-----KEEEKK----- 533  
Db 603 SKDPTSKPSDEP--EPKESDLTKTKKGFKTKGANYECTICGARYKKRDNYEAHKYYCS 660

Qy 534 -----DEEEKDEKDS-----KENTKEKID----- 555  
Db 661 ELQITKAHSVGAHEVETQAEPEWQSMHYMKLGATLELTLPLRKRKREKSLGDEEPPAF 720

Qy 556 -----GTAETEEREQATPR-----GRKTANSQGRKKRIETSM 589  
Db 721 ACPGPGSETAHRNPLJGTSKPAEASKAPSLEDPRASSGLPSQELGONQGRGECQPKKF 780

Qy 590 TNEAAAAASAAAAATEEP-----PPPL----- 611  
Db 781 TVIQTSSFEKSDPEQPSGLEEDKPPAQFSSPPPPAPHGRSAHLSQPLVROPNIQVPEI 840

Qy 612 -----PPPEPISTEPVETSRWTE-----EEMEVAKKGLVEHGRNMAAI 650  
Db 841 LVTEEPDRPDTEPEPPPEKPEKTEEFQWPQRTQTLAQPLAEKLPKKKRL-----RL 892

Qy 651 AKWYGTKSEACQKFNFYNYKRRHNDLLQHKOKTGRKPREERDVSQCESVASTVSAQE 710  
Db 893 AEMAQSGGESSFES-----SVPLSRSPSQESSISLS-----GSSRSASF 931

Qy 711 DEDIEASNEEENP-EDSEVEAV-----KPSEDSDPENATSRGTBPA 750  
Db 932 DREDHGKAFAPGPFSDTKSLTGLSHMLTVPSHHPHAREMRRSASEQSPNVHSSHMTETR 991

Qy 751 VELEPTTTETASTPSLAVPSTKPAEDSEVETQV-----NDSISAEATAEQMDVDOOEHS 805  
Db 992 SKSFDYDGLSLP-TGPSLAVPAAPPPPAAPPERRKCFVLROASLNRPPAEAEAVPRENR 1050

Qy 806 BEGSVCDPPPKADSDVDVVRVVENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQ 865  
Db 1051 AVRPAASKPTSKSVPOISVGTTCGGFS---GKSKQMDRPPGLGSS--PPYTEALQVFP 1106

Qy 866 INAQRPESQSDNSATCSADEVDGE--PERQRMFMDSKPSLNLPTGSLIVSSP----- 919  
Db 1107 LGTQLPPASLSFLSQLLPQEQOSSEFFP--TQAMAGLLSSPYSMPPLPSPSLFQAPPLPL 1165

Qy 920 ---LKNPDLDPLOHRAAVI---PPWVSCPTCNIPIGTPVSGYALYQRIKAMHESA 971  
Db 1166 QPTVLHPSQLHLPOLLPHAADIPFQPPSFLPMP- -PAPSTLSGY----- 1209

Qy 972 LLEQRQRQEOIDLECRSSTSPCGTSKSPNREWEVLQAPHQLITNLPEGVRLP- 1025  
Db 1210 FLPIQSQFALQDPEIESHLPPVKTSLPP---LATGPPCPSS--STEYSSDIQLPVPTFOA 1265

Qy 1026 -TTRPTRPPLIPSSKTTVAS-----EKPSF--IMGGSISQ-----GTP-GTY 1065  
Db 1266 TSPAPTSAPPLALPACPDAMVSLVVPVRIOTHPYSYGSAWYTLTSLIQLVQSPGSPASTA 1325

Qy 1066 LTSNQAASYTOET-----PKPS----- 1082  
Db 1326 LTKYEPPSKMTVCEADVAEAPGPPSSISKEQNRGYQTPYLVRPERKGTSLSSGILSL 1385

Qy 1083 -----VGSISGLPROQESAKSATLPVIKOEES----- 1111  
Db 1386 BGCSTASGSRKRVLSGASGLTMTQOQKR-----VKEEASKADEKLELVSTCSV 1438

Qy 1112 -----PRSQNSQPE-GLLVRAQHEGVVRGTAGA-----IOEGSITRGT----- 1148  
Db 1439 LTSTEDRKKTEKPHVGQGRSRREAETLSLSDSDVDPKELSPLSHSLSHGTAAGSEAL 1498

Qy 1149 -----PTSKISVESIPSLRGSITQGTALPQGTPTPEALVKGSIIRMPIDESSPKGREE 1203  
Db 1499 KEYAQPSKKAHRRGLPPM--SVKKEDPK-EQTDLPPLA---PPSSLSDTSKPAKLQ 1551

Qy 1204 AASKGHVIECKSGH-----ILSYDNINKAREGTR-----SPR-----T 1237  
Db 1552 BGTOSKVKVLPSPSLHTTNVSWCYLNYTKPNHIOHADRSSVYAGWCISLNPNLPGVST 1611

Qy 1238 AHEISLKRSYSEVGNIKQGMRESVPSAPLEGILICALPRGSPHSDLKERTVLGSGIM 1297  
Db 1612 KAALSLLRSKQKVS--KETYTMTAP--HPEAGRLVPSNRKPRMTEVHLPSVVVSPESQ 1666

Qy 1298 QGTPTATTE-----SPEDGLKYPKQIKRESPPIRAFEGAITKGPYDGTITTKEMGRSTH- 1352  
Db 1667 KDPARVEKEEKGKAEECT--PTSKRGEPARVKIPEGGYKSNEEY--IYVRGRGRGRV 1721

Qy 1353 -----EIPRODILTQESRKTPEVVQSTRPII-----EGSISOGTPIKFDNNSQ 1396  
Db 1722 CEEGIRCKFKPSLKKHIRTHTDV-----RPYCKHCHFAFKTKGNLTKHMKS----- 1770

Qy 1397 SAIKHNVKSLITGTSKLSRGMPPLEIV-----PENIKVVRGKVEDVKAG-- 1441  
Db 1771 ---AHSKKCQETGV-----LEELEAEBGTSDDLHQDSEGQGEAEVHEHOFSDLESDS 1821

Qy 1442 -----ETVRSRHTSVWSSGSPSLRSTLHEAPKAQLSPGIYDDTS---ARRTPVSYQ 1489  
Db 1822 DSOLDDEEEEEEEESODELSGP-----CSEAAPCLPPTLOENSSPVEGPAQDSTS 1875

Qy 1490 NTMRSGSPM---MMRTSDVTTPPNKSNHNRKSLITPTQRESIPAKSPVPGVDVWVSHSP 1546  
Db 1876 DEVPEGSGISBATHLTASSCSTPSRGTOGLPRLGLAPLEKMDMSAPSP--KATSP---RRP 1931

Qy 1547 FDPHHRGSTAGEVYWSHLPTQLDPAMPHPHALDPAANAALFQORQLSPPTPGVPSQVLYAM 1606

Db 1932 WSPSKAGSREPSVTRKHSLSLTNDSS---PQOCSPAREA---QASVTSTPG-PQ-----M 1978  
Qy 1607 ENTROTILINDYITSOQOVNLRPOVARGLSLSP---REOPGLGLP-----YPATRGI 1652  
Db 1979 GPGRD--LGPHLGLSP-----RPLSLR-LTPYPTIGREAPAGLERATDTGTPTSPTRWS 2030  
Qy 1653 IDLNMPTILVPHP---GGTSTPPMDR-----ITVPTQTTFPPRPVNSASMSFGHPH 1705  
Db 2031 LGAESPPOTVLPKQWALAGPCSPASAKSGGLGCPVPRALLQPVLPHTLLSRSP----- 2085  
Qy 1706 LAAASAERERERERERERERERERERERERERERERERERERERERERERERERERER 1765  
Db 2086 ETCTSAWRKTESRSP-----SAGPGLPFRPFS---APHDFHGLHPG-----RSEE 2128  
Qy 1766 TMLQORSVFGTNGTGVITPLDFTAQRLIMPAGPSPISQGLPASRYNTAADALAALV 1825  
Db 2129 NLFSLPLHSLGRAP--CPL-----IPIGQMVQARPGAQPTVLPGPCAAW 2176  
Qy 1826 DAAASAPOMDYKTKESKHEAARLEENLRSAASVSEOOQLEOKTLEVKRSVOCCLYTSS 1885  
Db 2177 SGFSGGSDLTGAREAQ-ERSRWSPT-ESPSASVPKAVSKFTLSSELEER-----T 2228  
Qy 1886 APFGSKP---OPHSSVYSEAGKDKGPPPKS 1913  
Db 2229 GRGPGRPDPWEPHRLKPLQDPWAHTAPAPHS 2259  
RESULT 21  
S15053  
hypothetical protein YCR033w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YCR592  
C:Species: Saccharomyces cerevisiae  
C:Date: 07-May-1993 #sequence.revision 07-May-1993 #text\_change 29-Oct-1999  
C:Accession: S15053; S19445; S40918  
R:Jia, Y.; Slonimski, P.P.; Herbert, C.J.  
Yeast 7, 413-424, 1991  
A:Title: The complete sequence of the unit YCR59, situated between CRY1 and MAT, reveals  
A:Reference number: S15052; MUID:91335897  
A:Accession: S15053  
A:Molecule type: DNA  
A:Residues: 1-1226 <JIA>  
A:Cross-references: EMBL:X59075; NID:g4805; PIDN:CAA41799.1; PID:g4807  
R:Herbert, C.J.; Jia, Y.; Slonimski, P.P.  
submitted to the Protein Sequence Database, March 1992  
A:Reference number: S19445  
A:Accession: S19445  
A:Molecule type: DNA  
A:Residues: 1-1226 <HER>  
A:Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42300.1; PID:e264374; PID:g190717  
R:Wicksteed, B.L.; Roberts, A.B.; Sagliocco, F.A.; Brown, A.J.P.  
Yeast 7, 761-772, 1991  
A:Title: The complete sequence of a 7.5 kb region of chromosome III from Saccharomyces  
A:Reference number: S40917; MUID:92133166  
A:Accession: S40918  
A:Molecule type: DNA  
A:Residues: 1-304, 'A', 306-374, 'S', 376-403, 'Q', 405-434, 'V', 436-441, 'N', 443-481, 'E', 483-511  
>  
A:Cross-references: EMBL:S78624; NID:g244237; PIDN:AAB21259.1; PID:g244239  
C:Genetics:  
A:Map position: 3R  
Query Match 3.1%; Score 390; DB 2; Length 1226;  
Best Local Similarity 20.2%; Pred. No. 6.2e-08;  
Matches 222; Conservative 181; Mismatches 368; Indels 326; Gaps 52;  
Qy 31 NTRHQEFVDPYDRSSHLEVSQASQLQOQQOQLRRRPSL-LSEFHPGS--DRPQ---- 83  
Db 247 NSIHORE---PFWKAN-----STILKTHSQS---SPSLHTKKFKDANKLDKPEASVK 294  
Qy 84 -----ERRT-SYE-----PFHPG----- 95  
| : | | |

Db 295 VETPSKDETKTISYHDNNFPKRSVSKPNAPLEPDNIKVGEDALGKKEVHKSGREIAKE 354  
Qy 96 -PSVP---DHDLSLESKPRLEQVSDSHFQVSAVLPVHPLPEGLRASADAKDPAGG 151  
Db 355 HPTVPRKHEDELEARAKVKNKI-----NIDGKODEIWT 389  
Qy 152 KHEAPSPISQPCDDQNASPSKKEELIQSDMDRVDREIAKVEQ-----QILKKKKQ 206  
Db 390 AKTVASA-----VEVSKEKHELTRSVRERKESPEIRDYRAYDPKALKTKDATK 437  
Qy 207 QOLEEEAAKPEP-----EKPVSPPPEQKHRSIVQIYDENRKK--AEAAKIFEGGLPK 260  
Db 438 LTVDDDNKSYEEPLEKVEGICFPLP---KAETRLWELKNOKRNIISKYVLLKKAIRNF 494  
Qy 261 VELPLYNQSDTKVYHENIKTNQVARKLILFFRRNHARKQREKQICORYDQLMEAWEK 320  
Db 495 SEYPFYAO---NKLIHQ---QATGLILTKIISKIKKEHLKK---INLKHDFDLOKKYEK 546  
Qy 321 KVD---RIENPRKAKESKREYEQFPEIRKORQOQERFORVGORGA--GLSATIAR 375  
Db 547 ECEILTSLSENLKREEIENRKEH-----ELMQKREEGIEKESLHPHSSSSSR 600  
Qy 376 SEHISEIIDGLSQEN-----NEKMRQLSVIPMMFDAQOR-RVKFINNGLMED- 426  
Db 601 RNRADFVDD--AEMENVLLQIDPNYKHQAAATIPPLIDPIRKYSYKFCFVNNLVTDK 658  
Qy 427 ---PMKVYKDRQPMNVWTDHEKEIFDKFIOHPKNFGLIASYL-ERKSVDPCLVLYLTK 482  
Db 659 KLNASRLKQAS--DNFTDHEHSLFLEGYLHPKPKFKISHYMGLKRSPECVLHYRTK 716  
Qy 483 KENYKALVRRNKGKRRNQIARPSQEEKVEKEDKAEKTEKKEEKK--DEEKEDE 540  
Db 717 KTVNYKQLLIDKNKKR---MSAAAKRRKRKESNDEEVEVESKESTNTIDKEESE 772  
Qy 541 KESKEN-----TKEKDIDGTAEETEREQATPRKKTAN--SQGRKGRITRSM 590  
Db 773 -NNAEENVQVLPVQSGSVKGDPLGTEKVE--NMIEKRGEEFAGELENAERVNDLKRHD 829  
Qy 591 NEAAAASAAAAATEEPPPLPPP-----PEPI-----STEPVET 625  
Db 830 EIGESKNSSVIETNNEVQIMAPKGGVNGYYPEETKELDFSLLENALQRKHKSAPEHKT 889  
Qy 626 SRWTEEMEYAKGLVEHGRNWAIAIAKMGVTKSAQCKNFYNYKRHRNLDNLQOHHKQK 685  
Db 890 SYNSVRESQLFPELLKEFGSOWSLSEKLGTKSTTMVRYNQRNAARNGWKLVDETDLK 949  
Qy 686 TSRKPREERDVSQCESVASTVSAQEDDEDIASNEENPEDESEAVK--PSEDSPE---- 739  
Db 950 -----RDGTSESV-----QOSQILIQPERPNINAYSNIPQORPALGYF 989  
Qy 740 -----NATSRGNTPEAVE-----LEPTTETAPSTSPSLAVPSTKPAED- 777  
Db 990 VGOPTHGNTSISIDGSIKPFDPDHRDFTSKISAPLTLPPLPRLPSIOFPSEMAEPT 1049  
Qy 778 -----ESVETQVNDTSISAEATAEQMDVDDQEHSAEBSGVCDDPPATKADSVDEVRY 828  
Db 1050 VTOLRNPLDHIOT-LADAASSVT-----NNQNFNSERNAI----- 1084  
Qy 829 PENHASKVEGDNTERDLDRASEKVEPRDEDLVVAQOINAEORPEQSDNDSSATCSADE- 887  
Db 1085 -----DIGRKTSTIS-----NLLNNSDRSMKSFQASRHEA 1116  
Qy 888 DVDGEPERQRMFMDSKPSLLNPTGSLVSPKLP-----NPLDLPQL--QHRAAVIP 938  
Db 1117 QLEDTPSMNNIVQEIKNITTPRSS-SISALLNPVNGSQSNPDGPRPLPFOHAISQGT 1175  
Qy 939 PMVSCPTCNIPITGPVS 955  
Db 1176 PTF---PLPAPRTSPIS 1189  
RESULT 22  
A41819



|    |      |                                                              |                                              |                                          |                                        |                             |       |           |       |      |
|----|------|--------------------------------------------------------------|----------------------------------------------|------------------------------------------|----------------------------------------|-----------------------------|-------|-----------|-------|------|
| Db | 254  | MetQTR                                                       | ---NPENSO---                                 | QeADPYfPQ                                | ----                                   | SPeEDPLIQEKEDPAHQHHTLTPTITG | 302   |           |       |      |
| Qy | 320  | KkVV                                                         | ---RIENNP---                                 | RRKAKESKTREY                             | YEQQPEPRKQREQOERTQVQORGAGLSATI         | 373                         |       |           |       |      |
| Db | 303  | KpVDIKIKITSEPIKIVESSIYEETPTQTGPFVFAKLFPSQ                    | -----                                        | QQPAETSETP                               | 355                                    |                             |       |           |       |      |
| Qy | 374  | ARSEHSEI                                                     | IIDGLSEQENNEKOMQRLSVIPPMFMDAEQRRVKFINNMGLMED | -----                                    | 426                                    |                             |       |           |       |      |
| Db | 356  | BEGESSGIDLESSVHPQDSE                                         | -----                                        | EIGPLTLQEDVSQ                            | -----                                  | 405                         |       |           |       |      |
| Qy | 427  | --PMKVYKDRQFMVMTDHEKEIFKD                                    | ----                                         | KFTQHPKNFGLIASYLER                       | --KSPVDCVLYY                           | 478                         |       |           |       |      |
| Db | 406  | EQPQLFESSEEGVSGSENQPEASVQPTVLPUYEQEILFGAGSSITVETLP           | -----                                        | IHE                                      | 462                                    |                             |       |           |       |      |
| Qy | 479  | YLTKKNNYKALVRNRYKRRG                                         | -----                                        | RNOIARPSQEEKVEKEEDKA                     | -----                                  | 522                         |       |           |       |      |
| Db | 463  | IQPTQNEY                                                     | -----                                        | GQLPNVTVPVDVALVTSEPVKTESFLAPQEPFVIA      | 510                                    |                             |       |           |       |      |
| Qy | 523  | -EKTEKKEEEKDEKEDKEDSKENTKEKDIDGTAETE                         | -                                            | EREQATPRGRKTANSQGR                       | 580                                    |                             |       |           |       |      |
| Db | 511  | LEYSNDVEPFVNEEPPVQAPETPGESQFQLEVPQAQATEYDEEFTKSATEQEQLAQFP   | 570                                          |                                          |                                        |                             |       |           |       |      |
| Qy | 581  | RKGRITSMWNEAAAAAATAEPP                                       | -----                                        | PPPLPPPEPI                               | -----                                  | 618                         |       |           |       |      |
| Db | 571  | ENDEVTWLPNSHYQAQHSILSNVTDOPLDLDTITTEKPMEMGTSVYVYDAAAAPEEVEFL | 630                                          |                                          |                                        |                             |       |           |       |      |
| Qy | 619  | -----STEPVETSRWTEEMEYAKKGLVHGRNWAIAK                         | ----                                         | MVG                                      | --TKSEAQCK                             | 663                         |       |           |       |      |
| Db | 631  | SDQQGVLSQSLEPILLYDSLSQOQYTTGISISRGGBFPPTQQETPHSGVMHTVEVAQPP  | 690                                          |                                          |                                        |                             |       |           |       |      |
| Qy | 664  | NFY                                                          | -----                                        | FNYKRRHN                                 | --LDNLIQHQKTSRKPREERDVSCESVASTVSAQEDIE | 715                         |       |           |       |      |
| Db | 691  | GHHEVTGPGLGHGVHSPALQNLITQY                                   | -----                                        | SIFPEKE                                  | ----                                   | QHSPVGLCV                   | 733   |           |       |      |
| Qy | 716  | ASNEENPDESEVAKVPSDESPENATSRNGTE                              | ----                                         | PAVELEPTTETAPSTSPSLAVPT                  | 772                                    |                             |       |           |       |      |
| Db | 734  | -----PGHLEEFVSPSPSOQENSAHWSTDMFLSPVDLQITFRSTQSKSYKTTI        | -----                                        | 783                                      |                                        |                             |       |           |       |      |
| Qy | 773  | KPAEDESVEITQVNDISIAETAE                                      | ----                                         | QMDVDOQEHSAEAG                           | -----                                  | SVCODPPATKAD                | 820   |           |       |      |
| Db | 784  | ---KHEDLATITPEPLEDGSILFPQEDLLQIDSTGQGEFSGHKTSTLSKPPVYSNTK    | 840                                          |                                          |                                        |                             |       |           |       |      |
| Qy | 821  | SVDEVVRPENHASKVEGDNTERKDLDRASEKVEPRDEDLVA                    | -----                                        | QQINAAQPEPQ                              | 874                                    |                             |       |           |       |      |
| Db | 841  | SSAQETWSETTY                                                 | -----                                        | NSKQVDLSSTHLKTSLEPPDYTMGLEPSLQQITQSSPKSL | 894                                    |                             |       |           |       |      |
| Qy | 875  | SDNDSSATCSADEVDGEPERQRMFMDSKPSL                              | -----                                        | LNPTGSILVSSPLKPN                         | 923                                    |                             |       |           |       |      |
| Db | 895  | HENPKS                                                       | -----                                        | FPV                                      | --LKPSAAQALKPKPEKLSPTNMIPHHPKSL        | 935                         |       |           |       |      |
| Qy | 924  | PLDLPLQLOHRAAVIPPMV                                          | -----                                        | SCTPCNIPICITPVSGYALYQRIKAMHES            | 970                                    |                             |       |           |       |      |
| Db | 936  | KNMWTTHPAHKMTVPQIQEDQGEYTISSNGSPQDLDEV                       | --                                           | LTLSGIIPEVKHIL                           | ----                                   | PK                          | 991   |           |       |      |
| Qy | 971  | ALLEEQROREQID                                                | -----                                        | LECRGSTPCGTSKSPNREWV                     | --LQAPAPOLITNLPEGVRLP                  | 1021                        |       |           |       |      |
| Db | 992  | RTVNPQYISQVKISHSOHVETQHPNSETTIVQPLDLFEALNQPTPE               | ----                                         | NFAQTQFDT                                | 1041                                   |                             |       |           |       |      |
| Qy | 1026 | TTRTPRPPPLIPSSKTTVAASEKPSFTMGSGISQGTPTGYLTSHNOASYTQETPKPSVGS | 1081                                         |                                          |                                        |                             |       |           |       |      |
| Db | 1049 | TTQMIGPPKEVI                                                 | -----                                        | QAPEHEHCT                                | 1071                                   |                             |       |           |       |      |
| Qy | 1086 | ISGLPQROQSAKSATILPYIKQEFSPRSQNSQEGLLVRAQHEGVVRGTAGIQBSIT     | 1141                                         |                                          |                                        |                             |       |           |       |      |
| Db | 1071 | IPIDI                                                        | -----                                        | QOAEYSITLPTV                             | --                                     | SFQPLDQE                    | ----- | LTTSSEAIR | ----- | 1101 |
| Qy | 1146 | RGTPTSKISVESI                                                | ----                                         | PSLRGSITQGTPALPQTGIPTEALVKG              | ----                                   | SISRMP                      | ----- | 1191      |       |      |
| Db | 1106 | --PHPTVPQOTIIVHPPEPLVHSEQVYQHPNTEALIQPLDLELITPQTPABEGL       | 1161                                         |                                          |                                        |                             |       |           |       |      |
| Qy | 1192 | ---IEDSSPEKGREAAASKGHV                                       | --                                           | IYEQKSGHILSYDNIKNAREGTSRPSRTHAISLKRSY    | 1241                                   |                             |       |           |       |      |

Db 1164 PQTLDSTQIIEPTTVVGVPIVEEVTQTSQDAEYPPSPVTSFQSL-DLELATSP 1222  
Qy 1248 ESVEGNIQGMRESVPASLEGICRALPRGSPHSDLKERTVLS-GSINQGTPTATTE 1306  
Db 1223 EATRESYHPSLLQQTIVNPPHEPLVLIHSEQVTHQHPNLTEATVQPLDPLDITIPQPTTE 1282  
Qy 1307 SFEDGLKPKQIKRES-----PIRAF-----EGAITKGKPYD-GIT 1342  
Db 1283 G-----ELPQLDSTQIIEPIKVVVALVYVQVSDQAEYTTSTVSFQLDQELT 1337  
Qy 1343 TIKEMGRSHE--IPRODIL-----TQSRKTPVEVQ-----STRPII 1378  
Db 1338 ITSEAIPEHPHTVPQQTITVHPHTKPLVLIHSEQTHQHPNTEVTVQPLDLELTMTQPTA 1397  
Qy 1379 EGSISQGTPIKFDNNGSOSALKHNVKSLITGPKSLRGMPLIIVPENIKVVERGKVEDV 1438  
Db 1398 EGELPQ-----TQDSTTQIIEPTTV-----VGVPI-----YEEV 1429  
Qy 1439 KAGETVRSRHTSVVSSGFSVLRSSTLHEAPKAQLSPGIYDDTSARTRPVSYQNTMSRGS-- 1496  
Db 1430 TVQTT--SQDAEYPPSPVTSFQSL-----DLGLTITPEPTEHFIQTQTV 1473  
Qy 1497 --PMNRTSDVTIPNKNSTNHERKSTLTPTORESIPAKSPVGVVDVPPVSHSPDPHHRGS 1554  
Db 1474 FPMY--TDVTLFQVSVQHLK-----PTEGIVQPLDLELTITP-QPTPEGE 1517  
Qy 1555 TAGEVYNSHLPTQDPAHPFHRALDPAAAYLFQRLSPGTPGYPXOYLAYMENTROTIL 1614  
Db 1518 LSQVQESTTQNK-EP-----HK--EVAPVPVYQAVTPT---FSQYQAEYKSLQPLDL 1567  
Qy 1615 NDYITTSQOMQVNLPRDVAR---GLSPREQLGLPYPATRGI--IDLTMPTTILVPHPG 1669  
Db 1568 ELTVTSEPTKAYHSTISKSLAINPQVHIQHPNAPATVQPLDLELTISSSLQPTAEG 1627  
Qy 1670 -----TSTPPMDRITYIPGQITPPRPNYSASPSGHPHLLAAASAEERERE 1717  
Db 1628 ELLYSMQETVTQISEPPKQVTVTPPEYEVAVPAPVQDAKYP--LSIVSLNSLDQELT 1685  
Qy 1718 REREKERERERAAASDLYLRGSEOPGRPGSHGYRSPSPSVRTO---ETMLQORPS 1773  
Db 1686 LSSELLGEAHLQITPDETMVLPKDRQGIYFDHDKHLNLTETVNPQPHLEHVQHQPT 1745  
Qy 1774 V-----FGTNGTSVITPLDPTAQLRIMPL----- 1799  
Db 1746 IEERSOSIQKTKTQITEPGKVVPLAQSESEVITPMPILKETAPPHPHSMALQSLDEKL 1805  
Qy 1800 ---AGGPSISQ---GLPASRYNTAADALAAVDAASAPQMDV----- 1836  
Db 1806 TIHSHSPGWTQOAHNLKESKGHTTG---KILLDYA--EPNIEELKHGHLFLKTTTEATT 1860  
Qy 1837 -----SKTKESKHEAARLEENLRSAVSEQQOLEQKLEVEKRSVOCLYTSAPPSG 1890  
Db 1861 ESENTNQMTSLKQVLTALFTQNKKSLMPALVSESDQSPPNMNSLQPLDQELTSSQPHG 1920  
Qy 1891 ----KQPHSSVYSEAKDKGP---PPKSRYEBELRTGKTTITTAANFIDVITRQIAS 1943  
Db 1921 WPHIPNTPKIIYLHAEPPTGPFVEPELFF--LKT-----TKSRPVQGTATQMAAS 1971  
Qy 1944 DKDARERGSSQSDSSLSLSHRYETPDSADAEIVSPASSPAPPEKLOTQYQPEVYKANAQAE 2003  
Db 1972 PKEMVSRAPENKEAVLS-----GPGE-----DQDESPSPNMSLQSLQDELTLSSQPH 2019  
Qy 2004 ----NDPTROEGPLHVRP----- 2019  
Db 2020 GWIHPPTNTHGKIYLHAEPPTGPFVEPPDLFLTKTKSPVETWLTFTDKSRKEMVSQS 2079  
Qy 2020 -----QOESPSP-----QQQLPSSQAEG-MQGVPRTHRLITLADHIC 2056  
Db 2080 KYEEAVLPVHGEGQESRSPPNMSLQSLQDELTLSSQPHGWPVHPNTHGKIYL--HYA 2137  
Qy 2057 QIITQDFARNQVSSQTQPOPTSTFQNSPALV-----STPVR--TKTSNRYSPESQASV 2110  
Db 2138 -----EPPTGPFVEPPDLFLTKTKSPVQGTATRMVKSPEEMVS-- 2177

Qy 2111 HHQPGSRVSPEN--LYDKSRGSRPGKSPERSH-----VSSEPYEPI-SPQVP 2156  
Db 2178 -----LDPENKEAVPPAQEGKGKSPSPNMSLQSLQDELTLSSQPHGWIHPHPRTP 2229  
Qy 2157 VVHEKQDSLILLSQGAEPABQRNDARSQGISYLPFF-----TKLE-----NTSPMVKSK 2208  
Db 2230 -----DKIYL--HYAEP-----PTGPFVEPPDLFLTKTKSPVQGTITTEMAKSP 2272  
Qy 2209 KOEI-----FRKLNSSG-GGSDMAAQAQGTETIFNLPAVTTSGVSSRGHS-FADPASNL 2261  
Db 2273 KEMVSQTPYKEAVLSGEGDQDESPPNPT---SLKSLDQEVAMSSQPHGVPHPKTP 2329  
Qy 2262 GLEDIIRKALMGSPDDKVEDHGVVMSOPMG-VVPGTANTSVVTSETRRREGDPS----- 2315  
Db 2330 G-----KIYLSI--EPPPGFVAPDLDLILVKTTKSKPAEWTPRRIDKL 2372  
Qy 2316 -----PHSGGYCKPKLISKNSRKSPIFGQGYLGTERPSSV-----SSVHS 2358  
Db 2373 LKEMVPHS-----PEYEAFFAHGEGQDESPPNMLQPLDQELTLSSQPHG 2421  
Qy 2359 EGDYHROTGP---WAWEDRPSSTGTSTOFFYNPLTMRMLSTP-----PTPACAFSAVNOA 2411  
Db 2422 WVPHPHTPGKIYLYHAEP--TGPFVEPPDLFLTKTKSPVQGSPEKTAISKEMVSQ 2479  
Qy 2412 APHOONRI-----WEREPAP 2426  
Db 2480 TPEYKEAVLSGEGDQDESPP 2501

RESULT 23  
T16871  
hypothetical protein T13H2.4 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000  
C;Accession: T16871  
R;WU, X.  
submitted to the EMBL Data Library, October 1995  
A;Description: The sequence of C. elegans cosmid T13H2.  
A;Reference number: Z18593  
A;Accession: T16871  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-2215 <WUX>  
A;Cross-references: EMBL:U39653; NID:g1049397; PID:g1049401; PIDN:AAB52495.1; GSPDB:G  
A;Experimental source: strain Bristol N2; clone T13H2  
C;Genetics:  
A;Gene: CESP:T13H2.4  
A;Map position: X  
A;Introns: 112/2; 136/1; 167/1; 196/1; 649/1; 696/1; 757/1; 850/1; 882/1; 985/1; 1046

Query Match 3.0%; Score 381.5; DB 2; Length 2215;  
Best Local Similarity 18.7%; Pred. No. 2.8e-07;  
Matches 444; Conservative 324; Mismatches 862; Indels 741; Gaps 105;

Qy 165 CGDONASPKLSKEELIQSMRDVREITAKVQOI-LKLKKKQOOLEEAAKPEPEKPV 223  
Db 112 CGHEPNL---RMNQE---DSSVVADESAGSPGELSGLQKVEVEDSWPQLRWEE 165  
Qy 224 SPPEVQKHRSIVQIYD---ENRKA-----EEAHKIFEGLPKVELFLYNQPSDTK 273  
Db 166 TPPTRRSTKMDLIGDVKPPGRIRATSYSCDIKRSGLISAHFASNDTVCVWVDQTS 225  
Qy 274 VYHENITNQVMRKLLIFPKRRNHARKQREKTCQYDQLMEAWKKVDRIENNRKA 333  
Db 226 IV---VASNOSTRDAL-----ASLCQR-----DSKVERMETAARISW 259  
Qy 334 KESKTREYVEKQPEIRKOREQOERFORVORGAGLSATIAARSEHSEIIDLGLSEENN 393  
Db 260 KLEDRVTEKTTURLTHEEYANARIQNNRNTQSLCSQPKRSRRVOGFID----- 311  
Qy 394 EKQMRQLSVIPPMFADAEQRVRVFIN-MNGLMEDPMKVYKDRQPMVWYWTQHEKEI--FKD 450

|    |      |                                                              |      |
|----|------|--------------------------------------------------------------|------|
| Db | 312  | -----PNNUTAEQAQMFYTYLNYTQDQFGKI-----GNLLHSEHKOLOQPLFN        | 354  |
| Qy | 451  | KFIQHPKFNGLIASYLERKSPVDCVLYLYLTKKN---ENYKALVRRNYKRRGRNQOIAI  | 507  |
| Db | 355  | K-----LHTSSVP---LHRLPQITFVN---STVRLEYSSSES-----              | 387  |
| Qy | 508  | PSOEKVEKEBEDKAEKTEKKEEKDEE-----EKDKEDSKENTKEKDIDG            | 556  |
| Db | 388  | PSNNNCCDAQNDRGHFGMSRPLENNTEHNATGISSPFLEKDSAPNPEDST-----      | 439  |
| Qy | 557  | TAEETEREQOATPRGRKTRANSOGRKGRITISMTNEAAAAAASAAAAATEEPPLPPPE   | 616  |
| Db | 440  | --PSVLNRDEASCTQKTRNS-----FEECVYNPALVLAAVNRVST-----PAT        | 481  |
| Qy | 617  | PISTE---PVEISRWTEEMEYAKGLVEHGRNMAAIKMYGKTEKSEACKNFYFNYYKRRH  | 673  |
| Db | 482  | PVINDVGPPWVFS-----ETANK-----AVVSIEGSMKIEOC-----H             | 514  |
| Qy | 674  | NLDNLLQOHKQTSRK---PREERDVQSCEVSATVSAQDEDEIA-----SNE          | 719  |
| Db | 515  | GPSNPQOLNNPETTRKSVSPLO-----RSMSTPASEATDEDLIAFLREPLALARAAE    | 567  |
| Qy | 720  | EENPESEVEAVKPSDESDENATSRGNETPAVEL-----EPTTATSTSPSLAVPSTKP    | 774  |
| Db | 568  | QEALRKKAEMPKLKKRSRAKNSAKRPVQTDVETVKLEKPAQSKSKNSRSKSTPS---    | 624  |
| Qy | 775  | AEDSEVETQVNDISAEATAEQMDVQOQSHSAEESGVCDPPPTKADSVDEVVRVPENHAS  | 834  |
| Db | 625  | -----SFANDOSSUSLEPLEM---RQHFVEGSELQKPSPELH-----PGTQGS        | 663  |
| Qy | 835  | KVEGDNTKERDLDRASEKVEPRDEDLVVAQINAQRP-----EPQSONDSSATC        | 883  |
| Db | 664  | SLRNFIRLDMPLQALQEV-PLG---AVAGQLRVQKALYKMKDMGNGIAPESASTSQTF   | 719  |
| Qy | 884  | SABEDVDGEPERQRMFPMDSKPSLLNPT-----GSI-LVSSPLKPNPLDLPQLOHRA    | 934  |
| Db | 720  | MODPAPPLPPKK-----AAPKRKKPTKABTEAAGNVSVSEKPKPRAPSKKKAQGV      | 772  |
| Qy | 935  | AVIPPMVSCV-----PCNPIGTPVSGYALYORHIKAMHESALL                  | 973  |
| Db | 773  | QQEQHQMISMTSAHQGTPEHPQSHOSTHQPLPSVANLLPLTGY-----             | 816  |
| Qy | 974  | EEQRQEQEIDLE-----CRSSTSPCGTSKSPKNREWEVLQ                     | 1009 |
| Db | 817  | QQOQOQOQRPSSQOHHGVNNNTTNGYVSDGNGSSSSSSSVSYSYSMQPTSDPP        | 876  |
| Qy | 1010 | APHOLITNLPUGVRLPTTRP-----TRPPPLPIPSKTT--                     | 1043 |
| Db | 877  | KKHQSLASMLTPIITGSSOPEATNGSSQIYTDKNQYHLYNNDNTYTPQPMYSNQTSN    | 936  |
| Qy | 1044 | -----VASEK-----PSFIMGSSISQCTPGTYLTSHNQASVTOETPKPSVGSISLG     | 1089 |
| Db | 937  | GYGGYEDVSNQFOQDPYPLPSVESQVSHQESQESNITHSSMPVTPISQOANNGSYNP    | 996  |
| Qy | 1090 | LPRQESAKSATLPIYKQ-----EEFSR-----SQNSQPEG---LLVR              | 1125 |
| Db | 997  | MPVTKDNRESNSSHILQDHEDLSVEQEAQYAPPSDNFFEMQNRSDQETNSAPLSMAR    | 1056 |
| Qy | 1126 | AQHEGVVRGTAGAIQEG---SITRTPTSKTSVESIPSLRSGITQGTPTALPQGTIPEAL  | 1182 |
| Db | 1057 | ASNASPFDELGYLDSGPGPSTSHQDPPADIELSDPPL---SI---TPQVAQEPVQT-TI  | 1109 |
| Qy | 1183 | VKGSISRMPTEDSSPEKGREEAASKGHVYBCKSHILLSYDNINKNAREGTRSPTAHEIS  | 1242 |
| Db | 1110 | MSNRVTKTPSSETHPO-----SYANVOYVPMNPSLPVSSHSES                  | 1147 |
| Qy | 1243 | LKRSYVESVGNITKQ-----GMSRESVPVS-----                          | 1266 |
| Db | 1148 | QVSTKLLPSNFKEDDIRHSTCPDNCNKCVCSEKSNADISMQPSFASITVTVPQSQORDQH | 1207 |
| Qy | 1267 | -----APLEGLIC---RALPRGSPHSDL---KERTVLSGSIQMGTPRATTESFED      | 1310 |

|   |   |      |                                                               |      |
|---|---|------|---------------------------------------------------------------|------|
| D | b | 1208 | LEQORGVEQLAANDDEQICVVKYKPLPHEISHTNLGPRROQSPSGVGVNGTSHGLTNGOPP | 126  |
| Q | y | 1311 | GLKYPOKIKRESPIRAFE---GAITKGPYDGIITIKEMGRSIEIPIRODILTQESRKT    | 1367 |
| D | b | 1268 | PLPPKRSRAKPKKNAFEDQMNASSGHAY-----QMQNS                        | 1303 |
| Q | y | 1368 | PEVVOSTRPIEGSISQ-----GTPIKFDNNSGQSAIKHNV--KSLITGPKSLSRGMP     | 1418 |
| D | b | 1304 | PEV-----SIGKLLKQOHLRPNQFNINSIQSOGLATLQOQSYMLSP-KLQRYSE        | 1352 |
| Q | y | 1419 | PLEI-----VPENIKVVERG--KYEDVKAGETVRSRHTSWSSGVSFLRSLTHE         | 1465 |
| D | b | 1353 | POTMKFQSHHQEAQVHEKKNHQOQSAKQEPHQIRQVNLQOVPTQOQSSGPA-HQSHIY    | 1411 |
| Q | y | 1466 | AP-KAOLSPGIYDDTSARRTPSVYQNTMRSGSPMNRRTSDVTIP-----PNKST--      | 1513 |
| D | b | 1412 | APMKQIK--LTSLOSNKTSQEHQSTTLQ-APVEEAPSD-QFPENLFQEQDSPDKVTEV    | 1466 |
| Q | y | 1514 | -NHERKSTL-----PPTQESIPAKS-PVPGVDVPVSHSPDPHHRGSTAGEV           | 1559 |
| D | b | 1467 | INNVA SVLSSPDEFQGOFTKTEKHLLDGTSKPTP-----KKOSHVDRRSRGSF--V     | 1518 |
| Q | y | 1560 | YWSHLPTQLDPAMPFFHRALDPAALAAIFORQLSPPTGYPSS---QYQLYAMENRTQITLN | 1615 |
| D | b | 1519 | HYVGQKQIMISQOQNRN-----AYQVOQOQPNASSLTMNEQYEV-VLDNTN--LQ       | 1567 |
| Q | y | 1616 | DYTSQOMQVNLRPDVARGLSPREQ-PLGLPYPATRGIIDLTMNPPPTILVP---HP----  | 1667 |
| D | b | 1568 | NPMSEQOMYNYQTSIVOOTSVEQOQPLQ-----QNGIQVTNQ-QTHRVYQHHVQHN      | 1621 |
| Q | y | 1668 | --GGTSTPMDRITYPGTQITFFPRPNYSASMSGHPHTL-----AAAAAERERER        | 1720 |
| D | b | 1622 | QRNGPTKAAPKR-----TPKPAVQSSKSVLHRAQMVDFAKTQPADQEDQTVQ          | 1672 |
| Q | y | 1721 | EKERERIAAASDLYLRPGSEQGRGPGSHGYVRSPSPSVRTQETMLQORPSVFOGTNG     | 1780 |
| D | b | 1673 | LQDHEHQ-----YNSQOQP-----EYOHQPLVQMQEQSNEQOQSVFQHQH-           | 1712 |
| Q | y | 1781 | TSVITPLDPTAQILRIMPLPAGGSPISOGPLASRYNTAADALAALYDAASAPOMDVSKTK  | 1840 |
| D | b | 1713 | -----HQAOQELAEQNG-----SMGMATQOQOCROQLOQLOQPPQ-QOHOIQOQY       | 1759 |
| Q | y | 1841 | ESKHEARLEENI---RSRSAAVSEQOOLEQRTLEV-EKRSVQCILYTSAPPSGKPOPHS   | 1896 |
| D | b | 1760 | QNVGPHSHIQONVAPQVRQOQPMQYAOQOQPMQMAQORSQMOEQYVOQGVQRTGATH     | 1819 |
| Q | y | 1897 | SVVYSEAGKDGPSPKRYEE---LRTGKTTITAA---NFDVIITRQIASDKRARER       | 1950 |
| D | b | 1820 | VLMTS-----PPNNHHESQANVLVRAPKQTIYKVSADPKPEDEAKTLIKLKDKLSD      | 1871 |
| Q | y | 1951 | GSQSSDSSSLSS-SHRYET-----PSDAIEVISPASSAPPQEKLOTY-----          | 1992 |
| D | b | 1872 | GQNAKPSVLRVKHIRETINIIALPSUIEELFDKADTPPDEYMEYMDFKITLH          | 1931 |
| Q | y | 1993 | ---QPEYVVKANOENDPTRYEGPLH-----HYRPQOESPQP-----                | 2027 |
| D | b | 1932 | MKDEPCVPEINVEDVLSTPLLEPALHGRPKRLAIAREGRTVTVTFFKKRKTASPEGDAK   | 1991 |
| Q | y | 2028 | -----QQLPSSQAEG--MCQVPRTHRLITLADHICQIITQDFARNQVSSQTPQ         | 2074 |
| D | b | 1992 | KRKPTNNHTNGMHTPDASPAQTKTSSQSSSS-----GMSQSLSTITTCMSSSS         | 2040 |
| Q | y | 2075 | QPPTSTFQNSPALSTPVRTKTSNRYSPESQAQSVHHORPGSRVSRPENLVKDSGRSPG    | 2134 |
| D | b | 2041 | SPASDFD-----ISTHLITPQEHSSPLTPTPIINHEYQVOYSSPVDIFSQSTSEPG      | 2094 |
| Q | y | 2135 | KSPERSHVSSEPYETSPQVPPVHKEQOSL                                 | 2165 |
| D | b | 2095 | -----PSSRPTRTGVHREAL                                          | 2111 |

RESULT  
T42761



Db 2173 LLPSTATVRAADGMISTIN-TPIAATLP-----ITTPQASVLRPMVRGG---MYRPG 2222  
Qy 1502 TSDVTIPNKNSTHERKSTLTPTQRESIPA-----KSPVGVDPVYVSHSPDFPHRGST 1555  
Db 2223 SGGVTAPLTSL-----TRVPMIAPRVLGAGLYRYPAPSRPIASTIP-----PA 2269  
Qy 1556 AGEVYWSHLPTQLDPAMPFHRALDPAAAYLFQRLSPGYPQSOYLAMENTRQTLN 1615  
Db 2270 EGPVYLG-----KASGAGGPP----- 2290  
Qy 1616 DYITSQOMVNLDPVARGLSRPOPLGLYPATRGIDLTNMPPTILVPHPGTSTPPM 1675  
Db 2291 -----RDELPAQ-GAREEPLS-----TTAPPVRIKEAPVAQAPAP- 2324  
Qy 1676 DRITYIPTQITFP--RPNYSASMP-----GHPTLAAASAE-RER-----ER 1718  
Db 2325 -----PPGOKPAGDAAGSGVGLGRPVMEKEASQEDRQKQOELLQLER 2371  
Qy 1719 ER-----EKERERERIA-----AASDLYLRP-----GSEQGRPGS 1750  
Db 2372 ERVELEKLRLQRLQELERERVELQRHREEQLLVQRELOLELOTIKHHVLOOQOERQAQ 2431  
Qy 1751 HGYVRSPPSVRTQETMLQQRPSVFQGTNGTSVTPDPTLAQLRIMPAGGPSISQGLP 1810  
Db 2432 FALQREOLAQRLOLEQTQLOOQLO-----QLEEQKQKQKAPFPATCEAPSRGPP 2483  
Qy 1811 ASRYNTAADA-----LAALVDAASAPQMDVSKTKESKHEAARLEENLRSAAVSE 1862  
Db 2484 PAATELANQGYWPLHTPIAVAGTEGP-----QOAREPV-----LHRLGPSASDMSL 2534  
Qy 1863 Q--QOLEQKLTLEVKRSVQCLYTSAPFSKPKQPHS-----SVVYSEAGKDKGPPPKSR 1914  
Db 2535 QTEQWEAGRSIGIKRHSMPRLRACPEPGDPSTVRRITADSSVQTDDEGEG-----R 2589  
Qy 1915 YEEELTRGKTTTAAAFIDVITRQIASDKD-----ARERGQSDSSSSLSH 1964  
Db 2590 Y--LLTRRRRTRSA--DCSVQ--TDEDNAEWEQPVRRRRSRLSRHSDSGSD-SKH 2639  
Qy 1965 RYETPSDA-----IEVSPASSPAPPOEKIQTQYQ-----EVKAN 2000  
Db 2640 EASASSAAAAAARMSVGIQTLSDCSVQTEP-EQLPRVSPATHIAATDPKVEIYRI 2698  
Qy 2001 QA-----ENDPTRQYEG-----PLHHRP-----QOBSPS 2025  
Db 2699 SAPEKTGRGESLAQTPEPDGAQVAGVQGLIGTAISPYPGLGIVTPGALGREKKPD 2758  
Qy 2026 P--QOQLPPSQAGMG-QVPRTHRLI-----TLADHICQIITQDPARNQ----- 2067  
Db 2759 PLEIGYQAHLPPELSQLVSRQPKSPQVLYSPVSPLSPH--RLLDTSFASSERLKAHV 2816  
Qy 2068 -----VSSQTPQOP--PTSTFQNSPALSVPVTRKTSNRYSPESQAGSVHHRP-G 2116  
Db 2817 SPQKQFTADSTLRQTLPRPKTLQRS-----LSDPKLSPTAESAKERSLYHQGGGLG 2872  
Qy 2117 SRVS--PENLVKRSRSPCKSPERSH-----VSSEPYEP-----ISPQVPPVYHE 2160  
Db 2873 SOVSAPENGLVRKVKRTLSPPEEAHLPLAGQVPSQVLAASLLQRLGAGTTPVAT-- 2930  
Qy 2161 QDSILLILLSQAGFAPQRNDARSPGISYLPSTFTKLENTSPVMKSKQBIFFKLNSG 2220  
Db 2931 -KASLLRELDRDLRVES-----TKLRKQABDEEKEIDAKLKYLE 2974  
Qy 2221 GGDSDMAAOPGTIFNLPAVTTSGSVSSRHSFA-----DPASNGLIEDIIRKALM 2272  
Db 2975 LG-----ITQRKESLAKDRVGRDYPPLRGLGEH-----RDYLS 3007  
Qy 2273 GSFDDKVEDHGVMSQPMGV---VPGTANTSVTSGETTRREE-----GDPSPH 2317  
Db 3008 DSELNQLRLOG--CTTAGQVVDYPASAANVATPSPGTAFCQPRFPAAQTYTAGSGGPT 3065  
Qy 2318 SGGVCKPKLISKSNRSKSPI--PGQGYLGTERTSPSVSSVHSGDYHROTP----- 2367

Db 3066 QNGFLAQAPTYTGPSTVPAPYPPGTSYPAEPQLSPQAFPHPTGHYAAPTMPMTQSAP 3125  
Qy 2368 -----GWAWEDRSSSTGCTOFFPNPLTMRMLSTPTPTPIACAPSANVQAAPHQONRIWE 2421  
Db 3126 FVQADSHAAHQKPRQSLADLEQKQVPTNYEIVISSPAVTVSSTPSETGYS----- 3175  
Qy 2422 REPAPLLSAQYE 2433  
Db 3176 --GPAVSSSYE 3184  
RESULT 25  
T16870  
hypothetical protein T13H2.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 02-Sep-2000  
C:Accession: T16870  
R:Wu, X.  
submitted to the EMBL Data Library, October 1995  
A:Description: The sequence of C. elegans cosmid T13H2.  
A:Reference number: Z18593  
A:Accession: T16870  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2447 <WUX>  
A:Cross-references: EMBL:U39653; NID:g1049397; PID:g1049400; PIDN:AAB52494.1; GSPDB:G  
A:Experimental source: strain Bristol N2; clone T13H2  
C:Genetics:  
A:Gene: CESP:T13H2.3  
A:Map position: X  
A:Introns: 18/3; 135/3; 176/3; 496/1; 693/1; 816/1; 1053/1; 1131/3; 1181/3; 1308/3; 1  
C:Superfamily: RING finger homology  
F:158-207/Domain: RING finger homology <RRN>

Query Match 3.0%; Score 376.5; DB 2; Length 2447;  
Best Local Similarity 19.1%; Pred. No. 4.9e-07;  
Matches 492; Conservative 309; Mismatches 892; Indels 881; Gaps 119;

Qy 98 PVDHDSLESKRPRLEQVSDSHFQVSAVLPVLVPLPEGLRASADAKKDPAGGKHEAPS 157  
Db 199 PTCRNLGKKR---ELOQDPDFDQLIYQVVESRIVGRMAENREHEKDVYFGRKYIEG 255  
Qy 158 SPISQPCGDDONASPSKLSKEELIQSMR-----VDREIAKVEQOILKLRKQO 207  
Db 256 GSDNKRKYIDPN---SKLKAPRLKSGAKKKIRWFHESDEGSGVRKWE-----SKGA 307  
Qy 208 QLEEE-----AAKPEPEKPVSPPPVQKHRSIVQI-IYDENRKKABEAHKI 253  
Db 308 PKEDDTNYLENDKEGTSVAARKEVLEEGEMDFPIEKSSDDEQTDLDDEESMLDSDFEI 367  
Qy 254 FEGLGPKVELP-----LYNQPSDTKVYHENIKTNQVMRKLILFKRRNHAR----- 300  
Db 368 SD--NEDVSKPSCSTSKTTNRSRSDSDSDNRDLQKRRKM--KRRKVPKTDGSDV 423  
Qy 301 -----KQEQKIQCR-----YDQLEAWERKVD 323  
Db 424 SNESFEDASGEVATKLKESKKKCGRPKKKFAPELIEGDIPTSEDSLTSSDEERD 483  
Qy 324 -----RTENPRKAKESKTYEYKOFPEIRKQREOQER-FQVGGQAGLS 370  
Db 484 NAADPYAFVQKEPNRDPDRDGHPEKDLY---NFDFMIDMNHQVDRKFEKDE----- 534  
Qy 371 ATARSHEILSEIDGLSEQENNEKQRLSVIPPMFDAQBRVKFINNGLMEDPMKV 430  
Db 535 -----THVISD---DSNEHESEDAEDRESSI-----DSE----- 561  
Qy 431 YKDRQFMNVWTDHEKEIFKDKFIQHPKFNGLIASYLERKSV-----DCVLYYLTCK 483  
Db 562 -----HKEIT--SKFLSH-----RQPLNPTSVDVDDCOV-ITVVKK 594  
Qy 484 NENYKALVRRNYGKRRGNQOIARPSQEEKVEEKDAEKTEKKEEKDEEKDEKED 543

|      |                                                               |      |
|------|---------------------------------------------------------------|------|
| 595  | DVQSAIT-----SKGETS-----PDSSSKTB-----EKPDVKVSEVSDDEMPHIT       | 638  |
| 544  | SKENT-----KERDKIDG-----TABETEERQATPRGRKTAN-----SQGR           | 580  |
| 639  | ADKGTDTFLNIMEHDEMYGGYLFPRGDTGISRPKVQAPGTNRLSMNVCEPAVLKEGK     | 698  |
| 581  | RK-----GRITRSMTWNEAAAAAASAAAAATEEPPLPP                        | 613  |
| 699  | KKLIVIPEDIYSDETVTLSDSEETSPSAEMEQSETSEAGPTIIRKTSGERETQSSS      | 758  |
| 614  | PPPISTPEVETR---WTEEMEYAKKGLVE-----HGRNWAIAKMWGTS              | 658  |
| 759  | PSEP-----STSRDRKMKRKLDRTRKRLADDSLDSDFDVSDGNELVATGKPI          | 808  |
| 659  | EAQCKNPFYNYKRRHNLNLQHQKQTSRKPREDRYSDQCESVAST                  | 705  |
| 809  | -IKHVFYDSANRMPSKSLDFTGRNAREIPMEIESRLAEQVAHEEYKTHRRQWVLE       | 867  |
| 706  | -----VSAQEDEDIASNEENPEDSEVEAVKPSDEPNATSRGNTPEPAVEL            | 753  |
| 868  | AVEAASKLNVVYDTEEBEIE---EETPEEVVYKASP-----TAPIATE              | 911  |
| 754  | EPTTETAPST-----SPSLAVPSTKPADESV-----ETQ-VNDSI                 | 788  |
| 912  | NPITSTAPFEEGVAMKETPIEIEFPDPEPCSSAAQARELIIERVGKEQOIIEDSLQEN    | 971  |
| 789  | ---SAETAQOMDV-DOQHSIAEGSVCDPPPATKADSVDVVVRVDPENHASKEVDNT--K   | 842  |
| 972  | RKPSSTTVAKESREAQPIEKDEMESEQOKDADNPTVEVD--KESEASSESDSKDFE      | 1030 |
| 843  | EROLDRASE--KVEPRDELVVAQOIN-----                               | 867  |
| 1031 | DETLDQAQSKTVKISLKHETVSDLEIEDFTYKFGEFVATAKAMIKRTIGEYVSTEFKL    | 1090 |
| 868  | -AQRPEQSDNSSATCSADENDVGEPERQRMFPMDSKPSLLNPTGSLVSSPLKPNPLD     | 926  |
| 1091 | VAOQPAVTDVLALGFCVNTDQEFSTIKE-----TGKRTKNPDDESVMK-----NFRE     | 1140 |
| 927  | LPQLOHRAAVIPWVNSCTPNICPTPVSGVALYORHI---KAMESALLFEORQOEI       | 983  |
| 1141 | SFAAKUR-----PVPRKPLTNIERMYIERAHVMVYKHYVDMEPLMKILIALQKQ---QI   | 1192 |
| 984  | DLECRSTSPCGTSKSPNREWEVLQAPH-----OLITNL---PEGVRLTPTRTPRPPPLI   | 1037 |
| 1193 | AATCAMLSPV-----TWTPPEAHAEQVQLLHLHLQNPSTLR-----PLI             | 1230 |
| 1038 | PSSKTTVAEKKPSFIMGGSISOGTPTVLTSHNQASYTQETPKPSVGSISIGLPROQESA   | 1097 |
| 1231 | -----NNFOFAL-----TLHKAQOQAIQOORAQQAQTKELA--ARQAEQA            | 1270 |
| 1098 | KSATL--PYIKOEFSRPR-----SONSOPEGLLV-----RAQHEGVVRGTAGAIQE      | 1141 |
| 1271 | RVEELARKRIAQEDAELAKRQGEQMSNVSGIPVSSDQNAQSSNAQQTGLIENOTTNNS    | 1330 |
| 1142 | GSITRGTPTSKIS-----VESIPLSGSTQGTTPALPQTGITEALVKGSTSRMP         | 1191 |
| 1331 | DSLTRPNTLADNSHLGSESOQPIVIESIQSSTSEALKESENTKOMPILTPASTVSSKSSAP | 1390 |
| 1192 | I--EDSSPKGREAASKGHVIYE--GRSGHILSYDNIKNAREGTRSPRAHEISLKRSY     | 1247 |
| 1391 | ATRRPSRCCSYDRPSSPVVIRERLGSOGALIN-----RPNRC-----               | 1430 |
| 1248 | ESVEGNIKOGMSWRSPVS-APLEGICRALPRGSPHSDLKERTVLVSGSIMGQTPRATTE   | 1306 |
| 1431 | -----NIDKRS--RSPISRAPVETV-----RINDHG--QNETILAGNI-----TH       | 1467 |
| 1307 | SFEDGLKYPKOIKRESPPITRAFEGAIKGGKPYDGTITTKMGRSIHEIPRODILITQESRK | 1366 |
| 1468 | TVE-----TTILEGTSI-----GQDSTI                                  | 1486 |
| 1367 | TPEVOSTRPIIEGTSOQGTPIKFNPN--SGOSATKHNVKSLITGPSKLSRGM-----     | 1417 |
| 1487 | RYDGECSQTYQIDKTD-----LDNSKNGTNVDEQSNVLKLRENDLREMLRYANRYH      | 1540 |

RESULT 26

RESULT  
T02345

102343 hypothetical protein KIAA0324 - human (fragment)

C:Species: Homo sapiens (man)

C; Species: homo sapiens (man)  
C; Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 05-Nov-1999

C;Accession: T02345

R; Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D

```

Query Match 3.08; Score 376; DB 2; Length 1791;
Best Local Similarity 19.9%; Pred. No. 3.5e-07;
Matches 396; Conservative 222; Mismatches 803; Indels 566; Gaps 75;

QY 681 QHKQTKRKPREERDVSCESVASTVSAQE-----DEDIEASNEEN 722
Db 50 QEKSKDSILVQSCPSLSLACGKVKSTPGESYFGVSSQLQKGQSTSPDHRSDTSSPEVR 109
QY 723 PEDSEVBANK-PSEDSPENATSRGNTPEAVLEPTTETAP-----STSPSL--- 767
Db 110 QSHSESPSLQSKQSTQPKGGRSRSS-----PWTELASRPIQRDGEFSASPMKSG 162
QY 768 AVPTKPAEDSVETQVNDSTLS-----AETAQMDVDOQOEHSAEGSVCDPPATK 818
Db 163 MSPQSRFQSDSSYPVIVSSNLGQSRLETAESKEKMLPPQ----BDATASPRQRKD 218
QY 819 ADSVDVEVRVPENHASKVEGDNTKERIDLRASEKVEPRDEDLVVAQINAORPEQSDND 878
Db 219 FSPFPVQDR-PE--SLVFNKDLTPTPRRSGA-----GSSPTEKEQNS 259
QY 879 SSATCSADEVDGEPERQRMFPMDSKPSLLNPTGSIL---VSSPLK----PNPLDLPQLQH 932
Db 260 ALPTSSQDEELMEVVEKSE-----EPAGQILSHLSSELKEMSTSNFESSPEVEE 308
QY 933 RAAV-----IPMVSCTCPNIGTPVSGYALVQVRIKAMHESALLEE 975
Db 309 RPAVSLTLQSQSOASLEAVEVPSMASW-----GGP-----HFSPEHKELSNP 353
QY 976 QROREQIDLECRSTSPCGTSKSPNREWEVLQ----PAPHOLITNLPEGVRLPTR--- 1028
Db 354 LRENSFGSPLEFRNS-GPLGTEMTMGFSSEYKEDLNGPFLNQLETDPSLDMKEQSTRSG 412
QY 1029 -----PTRPPPLIPSSKTTVASEKPSFTMG-----GSI 1057
Db 413 HSSSELSPDAVEKAGMSNQSISSPVLDAVPRTVSRKSSASSPEMKDGLPRTPSRNR 472
QY 1058 SQGTGTYLTSHNQASTQETPKPSVSGISLGLPRQOESAKSALPYIKQEEFSPRSON- 1116
Db 473 SGSSPGLRDGSGTPSRHSLSGSSPGMKDIPRTPSKRSECDSSPEPKALPQTPRPSRSP 532
QY 1117 SQPE--GLLVRAQHEGVVGTAGAIQESITRGTPTKISVESIPSLRGSTQGTPALPQ 1174
Db 533 SSPELNNKCLTQQRER--SGSESSVDQKTVAR--TPLQRS-----RSGSQELDVKPS 582
QY 1175 TGIPTALVKGSI-----SRMPTEDGSPEKGREAAKSHVI-YEGKSGHILSYDNIKN 1227
Db 583 AS-POERSESDSSPDSKAKTFTPLQRSSRSGSSSEVDSKSLSPRRSRG---SSPEVKD 638
QY 1228 -----AREGTRS-----PRTAHEISLKRYSYESVEGNIKOGMSRESFVSAPLEGL 1272
Db 639 KPRAAPRAQSGSDSSPEPKAPAPRALPRSRSGSSSKGRGSPESGSSSTESSPEHPKSR 698
QY 1273 ICRALPRGSPHDLKERTVLSGSIQGTGP-----RAATESFEDGLKYPKQIKR 1320
Db 699 TARGSRSSPEPKTKSRTPRRRSSRSPSELTKARLSRRSRSSASSSPETRSRTPPR-HR 757
QY 1321 ESPPIRAFEGA-----ITKGKPYDGIITIKEMGRSTHEIPRODILTQESRKTPEVQOS 1373

```

QY 2361 DYHROTGWAWEDRPSSTGTFYNNPLTMRMLSTPTPIACAPSAVNOQAPHQONRIW 2420  
Db 1670 -----SSSSSS-----SSSSSPAKPGQALPKPASPAPKPPG 1703  
QY 2421 EREAPL 2427  
Db 1704 ERSLLPV 1710

## RESULT 27

A47171  
chondroitin sulfate proteoglycan PG-M core protein - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
C:Accession: A47171  
R:Shinomura, T.; Nishida, Y.; Ito, K.; Kimata, K.  
J. Biol. Chem. 268, 14461-14469, 1993  
A:Title: cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan expressed during

A:Reference number: A47171; MUID:93300846

A:Accession: A47171  
A>Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-3562 <SH1>  
A:Cross-references: GB:D13542; NID:g391643; PIDN:BAA02742.1; PID:g391644  
A:Experimental source: stage 22-23 developing limb buds  
A:Note: sequence extracted from NCBI backbone (NCBI:134456, NCBI:134457)  
F:166-243/Domain: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lectin  
F:264-345/Domain: link protein repeat homology <LNK1>  
F:3258-3289/Domain: link protein repeat homology <LNK2>  
F:3296-3327/Domain: EGF homology <EGF1>  
F:3334-3454/Domain: C-type lectin homology <LCH>  
F:3461-3517/Domain: complement factor H repeat homology <PHD>

Query Match 3.08; Score 374.5; DB 2; Length 3562;  
Best Local Similarity 18.18; Pred. No. 9.5e-07;  
Matches 511; Conservative 379; Mismatches 1102; Indels 833; Gaps 118;

QY 9 NOGAFSTEO-SRYPPHSVQ-----YTPNTRHQO--EFAPVDYRSSHLEVSQSAL 56  
Db 781 SEGRFSEKFTHTPPVGMWMLQTDKQVYMTETSTKRIELDEDDISGMEPTSSPGQ 840  
QY 57 LQ-----QOQOOLRRRLSLSFHPGSDRQERRSYEPFHPGPS----- 97  
Db 841 LEYKHLGAPVSATDEKTSNETHAESDEEVSADFQTKGT-EVFTHTSSSLDLKFP 899  
QY 98 -----PVDHDS-----LEKRPRLQVSDSHFORVSAALPLVHLPLPEGLRA 139  
Db 900 TLSKIPEDSSATVKSFSSSSGTVLPTAVATVLETDHEADETSGLVNLMTFTSTPEGEQR 959  
QY 140 SADAKKPAFGKKEHAPSSPISGQGDQDQNASPSKLSKEELIOMDRVDREIAKVQOI 199  
Db 960 KA-TEKSPA-----TSAEDEVSTGTSKYTMTTEG-GQISSVTSAAEKESV 1002  
QY 200 LKAKKQOQ-----LLEAAKPPPEKPPVPPVE-----QKH 232  
Db 1003 AALQEREQPSVGLPEYKEPKFTDVTETIETVPQREGDTSLPVTVGSEDIGEMQVTDH 1062  
QY 233 RSIQIYIDENRKAERAKHIFEGLPKVELPLYNQPSDQTKVYHEN---IKTNQVMRKL 289  
Db 1063 TSFDSIIHTAATVSTRASEVF---PK-ELSTKDQDRELGTANGSTLPVTSVQMHQKT 1117  
QY 290 ILFFARRNHAKRQEQKICQRYDOLMEAKWEKKVDRI---ENNRPRKAKESKTRREYKEQF 346  
Db 1118 TAGFESQTTTQKHEHDEMSAYDMDYPATLSVPALMLTEYGVQSVGPVSTSTSLHGT 1177  
QY 347 PEIRKQEQERFORVQGRGAGLSATARSEHSEIIDLGLSEQENNEKQMRQLSVLPPM 406  
Db 1178 PKAETATDQEKITEAVPVFTGTOAKVYESK-----GTTTREDRDVGWSNVLPPH 1229

QY 407 MFDAEQRRVKFINMGLMEDPMKYVKDRQFMNVWTDHEKEIFKDKFKIQHPKNFGLIASYL 466  
Db 1230 TMLSSPSTAGSISLTLGASPS-----OTPEGSGISEELE 1264  
QY 467 ERKSVPDCLVLYLTKKNNENYKALVRRNYKRRGRNQOIARPSQOEKVEEKEDEAKTE 526  
Db 1265 EVKTPV-----FSSRATDKTTVI-----SDLTSSISAVDKIQPTASAKPFVSS 1308  
QY 527 KKEEKDDEEKDEKED-----SKENTYKDKIDGTAEETEERQATPRGRKKTANSQ 578  
Db 1309 KSPRIIPEDEEVTSSDIIVIDESISPSKASAEDDLTKGMVEP----- 1352  
QY 579 GRRKGRITRSMTNEAAAASAAAAATEEPPLPPPPPEPSTETPETSRTWTEEMEVAKK 638  
Db 1353 -----IDKEFYTSSTATAVAPTPTVMEATEALQOEVSPTSHPSDGTDIR-- 1400  
QY 639 GLVEHGRNWAIAARMVGTKSEAQCKNFYNYKRRHNLNLIQKHQKTKSRKPREERDVSQ 698  
Db 1401 -----LVVIQITGNDTHDPVNEFLDLFSRH-----ILPHAVDETHDTDAESAQTEP 1445  
QY 699 CESVASTVSA-----QEDEDIE-----ASN 718  
Db 1446 CTSDSVQDSSEYIILDPFFNFMDFEEDCENTDVTTPPALQFINGKQOVTSAKST 1505  
QY 719 EENPEDESEVAVKPSDSDPENATSRGNTPEPAVELEPTTETAPS-----TSPSL 767  
Db 1506 KAEARSDQIESVAHKNVTFQSIQNETNTFISETAGTMQPSKAGEVMGAFVTOPTA 1565  
QY 768 AVPTKPAEDSEVTQVNDISAEFTA--EQMDVDOQE-----HSAEGSVCDPPPATKADS 821  
Db 1566 DVAMLEVPYSGESEVTTDKYLEITSVYEQSPKKNKVTMMHGTESSTKD-----TK-- 1618  
QY 822 VDVEVRYPENHASKVEGDNKERDLDRASEKVEPRDELDVVAQOINQAQRPQSDNDSSA 881  
Db 1619 ---NLLITNESS---GDGSTESLSRS-----VFTEI-----L 1646  
QY 882 TCSADEVDGEPERQRMFMDSPKSLNLTGSIILVSSPLPNP-----LDLPQLQHRAAV 936  
Db 1647 TMSHED-----SEKISHTTSVPTILSVERSAVTAAPSADSDATVGDVKDL----- 1694  
QY 937 IPPMVSTPCNIPIGTPVSGYALYQRIKAMHESALLIEQRQOEI----- 983  
Db 1695 IPKGGTATPGN-----YKSTIKLDAEPFPFESNPEATSHHTKPKDMTASSFIVLE 1743  
QY 984 ---DLECRSTPCGTSKSPNREWEVLQAPAPHLITNLPEGVRLPTTRTPRPPPLIPSS 1040  
Db 1744 GSGDVEENSTILASAMTTETAETLSVQD-----TSLGSGTVLPTLS-----VTIS 1790  
QY 1041 KTTVASKEPFIIMGSGISQGTPTGYLTSHNOASYTQETPKPSVGSISLGLPROQESAKSA 1100  
Db 1791 EITPALPGGTRILYSTFDQSSEATVST-----NEVSELIMEQVVGSSVATEKKVEDEKEV 1845  
QY 1101 -TLPIYIKQE-----EFSP-----RSONSQPEGL-----LVRAQHEGVVGRGA 1136  
Db 1846 QTTVYISQEIETTDAGKSELDEFGSTTNEVTVSQEPTPLREIVPTITGMHSEIKKVT 1905  
QY 1137 GA-----IQEGS-----ITRGTPTSKISVESIP---SLRGSITQGTALPOTGIPTE 1180  
Db 1906 TPFLREKLFNEGSAEPADLFAGSPTRKVVSTDSPTDSSGSDIDVITESATLTSVPSR 1965  
QY 1181 ALV-----KGSIRMPIEDSSPEKGREAAKSHGVYIEKSGHILSYDNINKNAREGTR 1233  
Db 1966 SVIETQVTKHEGNINVISLKNITTEVEE-----HI---GTGGPVTSVSS-----TGSD 2012  
QY 1234 SPRTAHISILKRSVESYEGNIKQGMRESVPVAPLEGLICRALPRGSPHSDLKERVLS 1293  
Db 2013 GLTSESEVAIEMS--ENVFSTENQGEPTQE-----AVPTYTAPSDIKSRL--- 2055  
QY 1294 GSIOMGTPRATTESFEDGLKVPKQIKRESPPIRAFEGAITKGRPYDGTITTIKEMGRSIE 1353  
Db 2056 GSRREVTSHT-----PVIRT----- 2071  
QY 1354 IPRQDILTQESRKTPEVVQSTRPIIEGSIQGTPIKFDNNSGQSAIKHNKSLITGPKSL 1413



Db 2072 ---KDETAETVSPESV-----VNNSTLDMVTHGTIRAVAES 2110  
QY 1414 SRGMPLEIYP-ENIKVVERGKVEDKAGETVRSRHSVSVSSGPS--VLRS----- 1461  
Db 2111 KKGKGSFSAVSLGKILMIEHGSGEELK-----VDSSTTKLMSNGTEKLGSHSFDOGS 2166  
QY 1462 ---TLHEA- PKAQLSPGIYDDTSARTPVSY- QNTMSRGSPMMNRSTVDITPPNK---S 1512  
Db 2167 GEATLTSESTKASVSP-----TGKPEPQEQYGRKTVSMPSAVVHA---YTAEPNELVTS 2218  
QY 1513 TNHERKS--TLTPQRESIPAKSPVPGVDVPSVSHSPF--DPHHRGSTAGEVYWSHLPTOL 1568  
Db 2219 TEHDITSLQVTDTEMEKAANELT--VTSFATNLPLSESVHS-----WEDRPREI 2267  
QY 1569 DPAMPFHRALDPAAAA---YLFQRLS-----PTPCYPSQYQLYAMENTROTIL- 1614  
Db 2268 LP-----KATESGEATEDFFISTQANHEHVEFLSVPTTRPHEENKVAESDEKILLP 2322  
QY 1615 ---NDYIT-----SQQMVLNLRP--DVAR-GLSPREQPLGLPYPATR 1650  
Db 2323 FNDRVTVESAVERKYLSSPFTDEQEELVQNIFPTEDIPRLFTPKEE----- 2372  
QY 1651 GIIDLTMNPT--ILLVPHPGTSTPPMDRITYIPG-----TQITFPPRPYNSASMS 1700  
Db 2373 -----KPTNNELISDPLFSGGSGDEFTVIPSVESLAVKETNTLSPPWPFASVGP 2424  
QY 1701 ---CHPHTLAAAASAEERE--REREKERERERIAAASDLYLRPGSQPGRPGSHGVRS 1756  
Db 2425 KLSTDKQVFGSGTSDNAEINEITTAELTAYTSMATSSPALBEESSSHSNKDKD 2484  
QY 1757 PPSVRTQETMLQORSPVFGTNGTSVITPLDPTAQLRIMPLPAGGPSISQGLPASRYNT 1816  
Db 2485 ITHYFLVIEDPYNKEMDHRGNGTS-----RPLTPG---DVSLEESSHML 2528  
QY 1817 AADALAALVDAASAPOMDVSK-----TKESKHEAARLEENLRSGA--AVSEQ-----Q 1864  
Db 2529 TTDDVTPVSVILSETPYLEMGKSLATSATKMPSRVLPESSGEGWGVSDSFAPDTLTH 2588  
QY 1865 QLEQKTEVEKRSVQCLYTSAPPSGRQPQSHSVVYSEAGKDKGPPPKSRVEEELRTRGK 1924  
Db 2589 STAPSVMEVE-----ITASHIPG-----VYSEVMTHTVPGDGS-----Q 2623  
QY 1925 TTIT--AANFID--VIITQIADSKDARERGSSDSSSSLS----- 1962  
Db 2624 TVITGLASLFTKEEIVANRTAADPKGTSEELTSDTGMSLDILPVVDRRHVTLNVS 2683  
QY 1963 -----SHRYETPSDALEVISPASSAPPOE--KLQTYQPEVVKANAENDPTR----QY 2010  
Db 2684 GDITLIERLQIPSEKTTIIDMHSKMPEDILSVQTMPLNIVIRSTQVSDDNKMAEEDKY 2743  
QY 2011 EGPLHHRPOES-----PSPQOOLPSSQAGCMGOVPRTHRLITLADHICQLITODFA 2064  
Db 2744 DSIINFSTVEENSPGSGDNLSTISIQPSSESVTAGHGKLVKDLGSGVAMQFATLT 2803  
QY 2065 R-----NOVSSQTPQOOPPTSTFQNSPALSALVSTPVRTKSNRYSPESQASV 2110  
Db 2804 TTVLNLGILFPTVPSLVSPHMPHESKESEFEAKHIGTST-----TDDVTEPYTSANN- 2857  
QY 2111 HQRPGRSVENLVKSR-----GSRPGKSPERSHVSSEPEYIPSPQVPPVHKQ----- 2162  
Db 2858 -----QVITDQSKTMSISGFGMGQESG-DKPKMIPSLTDLTMEKALTLD 2905  
QY 2163 --DSLALLSORGAPAE-QRNDARSPCISYLV--PSFTKLENTSPMVKSKQEIFRKLN 2217  
Db 2906 TFDVSMVTTQSMQSHAVTSSSSSEKSTHYVMQTKASTEYEETDSV-----SLN 2955  
QY 2218 SSGGSDMAAAGPTEIFNLPAVTTSGSVSSRSHSPADPASNLGLEDIIRKALMGSD 2277  
Db 2956 SVSQNPASSVTVWLVNGVSKYKPEVIIPTSTSAKSDSDSDHSSD-----GTFKE 3003  
QY 2278 KVEDHGVVMSQPMGVPGTANTSVVTSGETRRE--EGDPSPHSGGVCKPKLISKNSRKS 2335

Db 3004 VSDMAATYKPTPTDLDTTVSSLLVFPESPESESISTESTPH-----FNKFVTERS 3054  
QY 2336 KSPIPGQGYLGTSPSSVS-----SVHSEGDYHROTPGWAMEDRPS-----STGSTQPP- 2384  
Db 3055 EETESSVNDLIIENATVSGDSPSIH---DY--PTAFWNGERTSTDVPKLSLSTIEVESS 3109  
QY 2385 ---YNP-----LPMRLMSSTPTPTPIACAPSAVNOAA--PHQONRIWEREPAPLLSAQYE 2433  
Db 3110 ERVKNPQESDRSTERERPRLLSSAPVSDSPNSIEGVFKPDQ-----EAVTMLTSSLE 3162  
QY 2434 TLDSDS 2438  
Db 3163 PLDRS 3167  
RESULT 28  
T20531  
hypothetical protein F07A11.6a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20531; T2776  
R:Palmer, S.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: Z19287  
A:Accession: T20531  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2526 <W1>  
A:Cross-references: EMBL:Z66511; PIDN:CAB54210.1; GSPDB:GN00020; CESP:F07A11.6a  
A:Experimental source: clone F07A11  
R:Gajadaty, S.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: Z20417  
A:Accession: T2776  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2526 <W12>  
A:Cross-references: EMBL:Z69904; PIDN:CAB54501.1; GSPDB:GN00020; CESP:F07A11.6a  
A:Experimental source: clone ZK20  
C:Genetics:  
A:Gene: CESP:F07A11.6a  
A:Map position: 2  
A:Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3  
Query Match 2.98; Score 372; DB 2; Length 2526;  
Best Local Similarity 18.28; Pred. No. 7.7e-07;  
Matches 467; Conservative 343; Mismatches 898; Indels 860; Gaps 115;  
QY 200 LKLLKKQOQLEEE-----AAKPPPEPEKVPSPPPVEQKHRSIVQIYIDENRKKAAEE----- 249  
Db 64 LRARSTRQPQEDSIEYKISAVIPEPTQN-SPIP-EFPSESTACVVYEQSGSTPERDLFE 121  
QY 250 ---AHKIFEGLGKVELPLYNQ--SDTKVYHENIKNQVMRKKLILFF---KRNHARK 301  
Db 122 LVKHKSRSGPIDIQLESTTEPGWKKARVHYRLDTGLKADKSLILGRPPKFRVYPT 181  
QY 302 QREOKICQRYDQLMEAWKVKVDRIENPRRKAKESKREYVEKOPPEIRKQREQOERFOR 361  
Db 182 SGEOK----HPQCHPSYSYAIPLKLGDHLLKASCSVHVPHLDHRSPD-----HYRRRFES 232  
QY 362 VQORGAGLSATIAARSEISEIIDGLSEQENNEKQRLSVIPPMFDAQORR---VKFI 418  
Db 233 YGQ---VIDVDMVKSNDKAPAVVQFTNIDDAQALQDNTNIPKMSYQSRSHRIIFYL 289  
QY 419 NMNGLMEDPMKVKY-----DROFMNVWTDHEKE--IFDKFTQHPKNFGLI 462  
Db 290 PIETNEEIMLIIRSLSDRIVDICVDWMDRSVAVITLDDMEPANLLKRMKLVGRNPF-- 347  
QY 463 ASYLERKSVDP-C---VLYYLLTKKNYKALVRNRYGKRGRNQOQIARPSQEKVEKE 518  
Db 348 ---EHRVAVDFCSRPNLYFINRKENKIEVAAKSSPTSSENDQGSPPSSSRDRONL 403



Db 2247 DSMGAEGSAFRILRSSTMGNSGSPSASGT-----TSPSTSSISS 2289

# RESULT 29

RBUAP

adenomatous polyposis coli protein - human

N;Alternate names: polyposis coli locus protein DP2.5; tumor suppressor APC

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1991 #sequence.revision 31-Dec-1991 #text\_change 21-Jul-2000

C;Accession: A3761; B39658; A44928; A49319; I54271

R;Kinzler, K.W.; Nilbert, M.C.; Su, L.K.; Vogelstein, B.; Bryan, T.M.; Levy, D.B.; Smith

chul, S.F.; Horii, A.; Ando, H.; Miyoshi, Y.; Miki, Y.; Nishisho, I.; Nakamura, Y.

Science 253, 661-665, 1991

A;Title: Identification of FAP locus genes from chromosome 5q21.

A;Reference number: A37261; MUID:91335210

A;Accession: A37261

A;Molecule type: mRNA

A;Residues: 1-2843 <KIN>

A;Cross-references: GB:M74088; NID:g182396; PIDN:AAA03586.1; PID:g182397

R;Joslyn, G.; Carlson, M.; Thilveris, A.; Albertsen, H.; Gelbert, L.; Samowitz, W.; Gro

arrington, J.; McPherson, J.; Wasmuth, J.; Le Paslier, D.; Abderrahim, H.; Cohen, D.; Le

Cell 66, 601-613, 1991

A;Title: Identification of deletion mutations and three new genes at the familial polypos

A;Reference number: A39658; MUID:91330307

A;Accession: B39658

A;Molecule type: DNA

A;Residues: 1-183, 'L', 185-969, 'N', 971-1308, 'G', 1310-1324, 'SS', 1326, 'HSTLE', 1332-1354, 'P'

A;Cross-references: GB:M73548; NID:g190163; PIDN:AAA60354.1; PID:g190164

R;Miki, Y.; Nishisho, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzler, K.W.; Vogelst

Cancer Res. 52, 643-645, 1992

A;Title: Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a c

A;Reference number: A44928; MUID:92119623

A;Accession: A44928

A;Molecule type: DNA

A;Residues: 1506-1525 <MIK>

A;Cross-references: GB:S78214; NID:g243541; PIDN:AB21145.1; PID:g243542

A;Note: sequence extracted from NCBI backbone (NCBIN:78214, NCBI:P:78218)

R;Spirio, L.; Olschwang, S.; Groden, J.; Robertson, M.; Samowitz, W.; Joslyn, G.; Gelber

Cell 75, 951-957, 1993

A;Title: Alleles of the APC gene: an attenuated form of familial polyposis.

A;Reference number: A49319; MUID:94073973

A;Accession: A49319

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 'G', 143-171, 'P', 173-179 <SPI>

A;Cross-references: GB:S67787; NID:g461061; PIDN:AAD13997.1; PID:g4261697

R;Lambertz, S.; Ballhausen, W.G.

Hum. Genet. 90, 650-652, 1993

A;Title: Identification of an alternative 5' untranslated region of the adenomatous poly

A;Reference number: I54271; MUID:93186137

A;Accession: I54271

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-4 <LAN>

A;Cross-references: GB:S56365; NID:g266243; PIDN:AAD14918.1; PID:g4262770

C;Genetics:

A;Gene: GDB:APC

A;Cross-references: GDB:119682; OMIM:175100

A;Map position: 5q21-5q22

A;Note: mutations of this gene can result in familial adenomatous polyposis or sporadic

C;Superfamily: adenomatous polyposis coli protein

C;Keywords: cancer; familial adenomatous polyposis; tumor suppressor

F;1-730/Domain: leucine-rich <NTD>

F;7-72/Region: coil #status predicted

F;185-227/Region: coil #status predicted

F;731-2832/Domain: serine-rich <CTD>

F;1131-1156/Region: acidic

F;1558-1577/Region: acidic

F;1866-1893/Region: highly charged

Query Match

Best Local Similarity 2.9%; Score 365; DB 1; Length 2843;

19.0%; Pred. No. 1.7e-06;

Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

QY 347 PEIRKQEQOE-RFORVGORGAGLSATIRSEH-----EISEIIDGLSEQENNEK 395  
 Db 981 PSIESYEDDESCKFCYQYPADLAHKIHSANHMDNDGELDTPLNYSKYSDQOLNSGR 1040  
 QY 396 QMRQLSVIPPMFDAQRRVKFINNGLMEDPMKVYKDRQPMYVTHDE--KEIFKDFI 453  
 Db 1041 Q-----SPQNERWARPKHIEDEIKQSEQRQSRQNSTTVPVYTESTDDKHL 1087  
 QY 454 QHPKNFGLIASYLERKSVDPDCLVYYITKKNNYKALVRRNYGRRGRNQOIAJPQSEK 513  
 Db 1088 KFOPHFG-----QQECVSPY-----RSRGANGSETNRVGNHGINQNVSLQCE 1132  
 QY 514 VEKEDKAEKTEKK--EEKKDEEKDEKEDSKENTKEDKIDGTAEETEEREQATPRG 571  
 Db 1133 -DYEDDKPTNYSERYSEEEQHEEERTNYSIKYN-EKRRHVDPDIDYSUKYATDIPSS 1190  
 QY 572 R-----KTANSQGRRKGRITRSMTNEAAAS-----AAAAAATPEPPPLPPP 614  
 Db 1191 QKQSFSEKSSGSGSSKTEHMSSESSTSTPSSNAKRNQLHPSSAQSRSGQPKAATCK 1250  
 QY 615 PEPISTPEVET-----SRTEEMEVAKKGLVEHGRN-----WAAIAKM 653  
 Db 1251 VSSINQETIQTICVEDTPICFSCSSLSLSSAEDGICNQTQEADSANTLQIAEIK 1310  
 QY 654 VGTKSE-----AOCKNFYNYKRRHNLNLOOHK-----OQTSR 688  
 Db 1311 IGTSAEDPVSEVPVAVCHPRTKSRRLQSSLSSESRHKAVFSSGAKSPKSGAQTTPK 1370  
 QY 689 KPREE-----RDVSOCEVASTVSAQDEDEIASNEENEPDESEVAVKPSE--DSPEN 740  
 Db 1371 SPPEHYVQETPLMFESRCTSV--SLDSFESRSIASSVQSEPCSGMVSGIISPDLPSPGQ 1429  
 QY 741 A--TSRNTPEPAVELEPTTETAPSTPSLAVSTPKPAEDESVEVQVNDISAEATQMDV 798  
 Db 1430 TMPFSRSKTPPP--PPQATQTRREVKNKAPTAEKRESGPKQAAVAAVQ--RVQVLPDA 1485  
 QY 799 DQOQSHAE-----GSGV-----DPPATKADSVDEVVR-----VPEN-HASKVEGDN 840  
 Db 1486 DTLHFATSTPDGFCSSLSLSALSDPEFTQK----DVELRIMPVQENDNGNETESEQ 1541  
 QY 841 TKE--RDLDRASEKVEPRDEDLVVAQIINAQRPQSDNDSSATCSADEQVD----- 890  
 Db 1542 PKESNQEKEAEKTIQSEKDL-----DQSD-----DDOIELEECIIS 1581  
 QY 891 --GEPERQRMFMDSKPSSLNPTGSLVSSPLKPNPLD---LP-----QLQHRRAVIP 938  
 Db 1582 AMPTKSSRKAKKPAQTAASKLPPPVAR-----KPSQLPVYKLLPSQNRQLQPKHVSFT 1634  
 QY 939 ---PMVSC---TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
 Db 1635 GDDMPRYVCEGTPINSTATSLDITIEPPNELAAGEVGRGAQSGEPEKRTITFEG 1694  
 QY 971 AL-----LEEQRQREQIDLECRSSSTPCGTSKSPNREWEVL----- 1007  
 Db 1695 RSTDEAOGKTSSVTIPELDDNKAEGDILAECINSAMPKSKHFKVRKKIMDQVOOAS 1754  
 QY 1008 --QAPQLITNLPEGVRLPTTRTPPPPLI-----PS 1039  
 Db 1755 ASSAPNK---NQLDGKKKKPTSPVKPIQONTYRTRVRKNADSKNNLNAERFVSDNKS 1811  
 QY 1040 SKTTVASEKPSF-----IMGSGTS-----QGTPTGYLTSHQOASYQTETPK 1080  
 Db 1812 KKNLKNNSKDFNDKLPNNEDRVGRSFAFDSPPHYTIEGTP--YCFSRND----- 1860  
 QY 1081 PSVGISLGLPROQESAKSATLPYIKQEEFSPRSQNSQOPEGLLVYRAQHEGVYRTAGAIQ 1140  
 Db 1861 -SLSSLDFFDDVDSLREKAELRAKENKESEAKVTSHTE--LTSNQOSA---NKTQAI 1914  
 QY 1141 EGSITRGTPTSKISVESIPSLRGSITQGTALPQTGTPTEALVK-----GS 1186  
 Db 1915 KQPINRGOPKPILOKQS-----TFPOSSKIDIPRGAATDEKLQNFATIENTPVCFSHNS 1968



|    |      |                                                                |      |
|----|------|----------------------------------------------------------------|------|
| Qy | 1071 | QASVTQETPKPSVGSISLGLPQQSASAKSATLPYIKQEESPRSONSGOPEGLLVRAHQEG   | 1130 |
| Db | 687  | VKSKENTPKRAENDTINTLPGGEGPEDNAEK---RNEAVSPNDTSE-----IKQDLEN     | 739  |
| Qy | 1131 | VVRGTAGAOEGSITRGTTYSKISVESIPSLRGSITOGTTPALPQT-----CPT          | 1179 |
| Db | 740  | LENGASGPNNVQVDAAEQDPTDPTDEVTTSKISENNPKAPOTEDDNDATFTEAGLET      | 799  |
| Qy | 1180 | -----EALVAGSIRMPIED---SSPEKGREEAASKGHVIECKSGSHIL               | 1220 |
| Db | 800  | TEAFGDAEHVSYLDANIEKLVAWADEPLPVDELVSIEERPEVAPAES---TGEDEDIF     | 855  |
| Qy | 1221 | SYONIKNAREGTRSPRTAHE-ISLKRSEYSEVGNIKQCMRESRSPVSAPLEGLICRALPR   | 1279 |
| Db | 856  | RRDRRTVSLTGTGDQNAPIQVIFVGDDGDNPDANADQ-----ER                   | 894  |
| Qy | 1280 | GSPHSDLKERTVLSSGIMOGTTPRAYTESFEDGLKYPKQIKRESP-----PIRA         | 1327 |
| Db | 895  | TSHNEILIESD-----KESEBAITKNEEDVDQDP---IQSEELPTSOEGESSIGNKIVA    | 945  |
| Qy | 1328 | FEGAITKKG---PYDGIIT-----IKEMGRSHEIIPROD-----ILT                | 1361 |
| Db | 946  | VVGSVLLGGAVIPYGVLASNENADAHREVEETGDSFDRPEETFVSKLTSVMENILP       | 1005 |
| Qy | 1362 | QESRKYPEVVQSTRPIIEGSIQGTPIKPDNNNGOSAKIHN-VKSLITGPSKLSRGMPLP    | 1420 |
| Db | 1006 | STNDENPEAVSMVENYLPVNTGLDESKEDNDPAPTAHAHSGEKNLRNDKTTDTRERDPE    | 1065 |
| Qy | 1421 | EIVPENI-KVBERGYEDVKAGETVRSRHSYVSSGGSVLRSLTHEAPKAO-LSPGIYD      | 1477 |
| Db | 1066 | E---ETILNKLVE-----NALPGVTGSGTFTEV-----SAPDAQELDETVVD           | 1103 |
| Qy | 1478 | DTSARRTPSVYQNTMSRGSPMMNRSTDVTIPNKNSTNHERKSTLTPTQRESIPAKSPVPG   | 1537 |
| Db | 1104 | HAG-----QNDTSEVEDAPEKSAGGTVEIKETFSMIESLIPVQAP---1142           |      |
| Qy | 1538 | VDPVWSHSPDPHHRGSTAGEYVWSHLPTQDLPAMPFHRAIDLPDAAAAYLFQORLSPPTPGY | 1597 |
| Db | 1143 | TOPENAESHIQDTGG---ASEI-----KDDQNQP-----EESFAHQGFVEYSAPDQ       | 1188 |
| Qy | 1598 | PS-OYOLYA-MENTROTILNDYITSOQMOVLNRPDVARGLSPREQLPGLPYPATRGIIDL   | 1655 |
| Db | 1189 | ESAEFQLEAKKODKQETIENSEDAKTVMEKLVSLVENILPVSEAVLPSDSTVTKNSEDK    | 1248 |
| Qy | 1656 | TNMPPTILVPHPGGTSTPPMDRITVPGTQITFFPRPY-----NSASMSPGH            | 1702 |
| Db | 1249 | KELETQELSSKEIKTSGQP----EYVPTSEAFVSDPELFORVKRASTSEPKTQKTEPHA    | 1304 |
| Qy | 1703 | PTHLAAAAGARERE-----REREKERERERIAAASSDLYLRPGEOPGRPGSHGY----     | 1753 |
| Db | 1305 | PIFVQSGSTEDDEQSIANVIDELVHEDDEKKVPEVTANISVSASENIDDDSTANAVPKTE   | 1364 |
| Qy | 1754 | VRSPSPSVRTOETHLQORPSVFOGTNGTSVITPLDPTAQLRIMPLPAGGSPISQGLPASR   | 1813 |
| Db | 1365 | VSQEQLOVATFVEFEESAPE----EESAAIPEVOEPLKEVQ-----PDLQNSQSPAH      | 1414 |
| Qy | 1814 | YNTAADALAALVDAAASAPQMD-----VSKTESKHEAARLEN-----1852            |      |
| Db | 1415 | -----KIDLHNTPKHEDYNDYVFPFGTESSEESQKADGNQENDEEDDVVAELN          | 1465 |
| Qy | 1853 | -----LRSRSAAVSE-----QQOLEPKTKLEV-----EKRSVQC                   | 1880 |
| Db | 1466 | FHPIRQWRDETVISLQSLKSLVAEVCITDVSADVNEQDEESTLILKVVPSEPSLLEL      | 1525 |
| Qy | 1881 | LYTSSAPPGGKQPPSHSVVYVSAGKDKGPPKPSRYEELTRGKTTITAANFIDVITRQ      | 1940 |
| Db | 1526 | DFTNDPKRIVHPITLME-----PATMYLEEM-----VEMIIADAVKSEME             | 1567 |
| Qy | 1941 | IASDKDARERGSSDSSSLSSHRYETPSDAIEVISPASSPAPPOKLOTQYQEVVKAN       | 2000 |
| Db | 1568 | VVTEISEISMAPQVSESTCPIP-----EPLADLKLPVEDDEBK---TPPEPVVPG        | 1614 |

RESULT 31

RESULTS 31  
T26215

140213  
hypothetical protein W06A7.3a - *Caenorhabditis elegans*

C: Species: *Caenorhabditis elegans*

C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C: Accession: T26215

R: Ainscough, R.

R; Almscough, R. submitted to the EMBL Data Library. August 1996

A:Reference number: Z20173

A: Accession: T26215

A;Accession: T26215  
A-Status: preliminary, translated from CR/EMBL/DBP

A: Status: preliminary; t:  
A: Molecule type: DNA

A;Molecule type: DNA  
A:Residues: 1-2607 >WT1>

A;Residues: 1-2607 <WIL>  
A:Cross-references: EMP1:778066: PIDN:CAP01522 3: CSDDB:CN000023: CESP:W06A7.2a

A; Cross-references: EMBL: Z78066; F  
X: Experimental source: clone W06A7

A; Experiment

C;Genetics:

A;Gene: CESP:W06A

A;Map position: 5

| Query Match | 2.98: | Score 361: | DB 2: | Length 2607: |
|-------------|-------|------------|-------|--------------|
|-------------|-------|------------|-------|--------------|

Query Match 2.9%; Score 361; DB 2; Length 2607;  
Best Local Similarity 18.1%; Pred. No. 2.1e-06;

|                       |                   |                    |
|-----------------------|-------------------|--------------------|
| BEST LOCAL SIMILARITY | 18.1%;            | Pred. NO. 2.1e-06; |
| Matches 428:          | Conservative 320: | Mismatches 937:    |
|                       | Indels 686:       | Gaps 95:           |

07 005 006 007 008 009 010 011 012 013 014 015 016 017 018 019 020 021 022 023 024 025 026 027 028 029 030 031 032 033 034 035 036 037 038 039 040 041 042 043 044 045 046 047 048 049 050 051 052 053 054 055

QY 509 SQEEKVEEEDKAEKTEKKE-----EKKDEEEKDEKEDSKENTKEKDIDGTA 558

Db 25 SDQESSENKEPDASENVEETEPVSSAKVIALESSGDCEEKNIKISANENSVEPDGADKPA 84

|    |      |                                                                |      |
|----|------|----------------------------------------------------------------|------|
| Qy | 1478 | DTSARRPVSQYNTMRGSGPMNRSTSDVTIPPNKTNHERKSTLTPQRSIPAKSPVPG       | 1533 |
| Db | 1104 | HAG-----QNDTSEVEDAPEKSAGGVTEKTSMTESTILPVQAP----                | 1142 |
| Qy | 1538 | VDPVVSHSPDPPHRCSTAGEYWWSHLPTQLDPAMPFHRLDPAAXAYLFLQRQLSPTPGY    | 1597 |
| Db | 1143 | TQPENAESHIQDTGG--ASEI-----KDQNOP-----EFSAHQCKFEVSAEPDQ         | 1188 |
| Qy | 1598 | PS-QYOIYA-MENTROTILNDVITSOQMVOYNLRPDVARGLSPREOPLGLPYPATRGIDL   | 1655 |
| Db | 1189 | ESAEFQLEAKDOOKETIENSEDAKETWVKLSLVENIILPFAVLPUSTVTKNSEDK        | 1248 |
| Qy | 1656 | TNMPPTILVPHPGGTSTPPMDRITYIGTQTITFPRPY-----NSASMSGPH            | 1702 |
| Db | 1249 | KELETQELSSKEIKSQP-----EYVPTSEAFVSDPEIFQVRKRASSTEPTKQKTEPHA     | 1304 |
| Qy | 1703 | PTHAAAAAASERRE-----REREKERERTAAASSDLYLRPOSEQPRGRSGHY           | 1753 |
| Db | 1305 | PITVGOSTEDDEBQSIA NYDELVLVHEDEKKVPEYTANISVASENIDDSTTANA VPKE   | 1364 |
| Qy | 1754 | VRSPSPSVRTQETMLQORPSVFQGTNGTSVITPDPDTAQLRMPLPAGGSIISOGLPASR    | 1813 |
| Db | 1365 | VSSEQLOVATEFELGSAPE---EESAIPA EVQPLEKEVEQ-----PDLSONSPAPH      | 1414 |
| Qy | 1814 | YNTAADALALVDAAAASAPQMD-----VSKTESKHEAARLEN-----                | 1852 |
| Db | 1415 | -----KIIDLHFNIKDHEDYNDYVPGTSESSESKADGNQENQEEDVYVAELN           | 1465 |
| Qy | 1853 | -----LKSRSAAYSE-----QQOLEQKTELEV-----EKRSVQC                   | 1880 |
| Db | 1466 | FPIROWRDEDVISLOSLASLVAEVCITDVADSNVQDEESTKLTKLVVSPSPSLLEL       | 1525 |
| Qy | 1881 | LYTSSAFFPSGKPQHSSSVVSYSEAGDKGPPPKSRVEEELTRGKTTITAANFDVIITRQ    | 1940 |
| Db | 1526 | DFNDPKV IHVP I PLME-----PATMYLEEM-----VENIIDAVKVESEME          | 1567 |
| Qy | 1941 | IASKDARERGSSDSSSLSSHRYETPSDAIEVISPASSAPPQPEKLOTYOPEVVKAN       | 2000 |
| Db | 1568 | VVTEISEMANPOVSESTCIP-----EPLADLKL PVEDDEK--TPEPEVPVPG          | 1614 |
| Qy | 2001 | QAEND--PTRQYEGPLHHVRRPOEESPSPQOQLP----PSSQAEQMG-----QVPRTH-    | 2046 |
| Db | 1615 | QVOERIPIEVEQAPT---IPORPPRAPKSELPAKVAKPLDDSKSRVRFAPLNKLGRTYS    | 1671 |
| Qy | 2047 | ----RLITLADHICOIITODFARNQVSSOTPOQPTSTF-----QNSPSALV---STPVR    | 2094 |
| Db | 1672 | EEQOKELVESLERPLTIHQ-----QKPPEKTEDIGALSPLSPNTLA EYEVPM          | 1722 |
| Qy | 2095 | TKTSNRYSPESQAQSVHH-----QRPGSRVSPENLVDRSGSRGPKSPERSHVSSPEYP     | 2149 |
| Db | 1723 | DMQSVPHSQOEKEETEALSEIIIEBPQAMKEVPVE-----SAPEKNDESLEAPEI        | 1774 |
| Qy | 2150 | ISPQVPVNVHEKO-----DSLLLLSQRGA-----EPAEQR                       | 2179 |
| Db | 1775 | INEPIRVIVETKIMGPGKSLNEDDDDGSECILSDIGLSERTIQRFNTSIDDPSIRR       | 1834 |
| Qy | 2180 | NDARSPCSISYLPSFTFKLENTSPMVKSQKEIFRKLNSSGGD-----SDMAAAQPGTE     | 2234 |
| Db | 1835 | DSFSSISSFGDRQKFATIENTRODLLPQGSVSQYLRSPNPSSQQLLVLTNLMSDPSDL     | 1894 |
| Qy | 2235 | IFNLPAV-----TTSQGVSSRR-----HSFADPASNGLGEDIRKAL                 | 2271 |
| Db | 1895 | SPNAPVGFENTA QFLEKLOQEDRPSAEGSIDSSGFKEKVDHEGLDEF AAPVHD PMQXSV | 1954 |
| Qy | 2272 | MGSF--DDK---VEDHG VVYNSQPMGVPGTANTSVVTSGETRRREE-----GD PSP     | 2316 |
| Db | 1955 | FGSIGSDMKPKGSDQDGFV-----IERNEANEATLAKNKMSHHNDVIEKNYFDNAP       | 2009 |
| Qy | 2317 | HSGGVCKP-----KLIS---KSNRSKSKSPIPGOGYLGTE-----RPS SVSS          | 2355 |
| Db | 2010 | TAALLIESPIAEARKULQDVAIVESAYEYKQAVDSGEIGRELLDNVBQKLEQVKEP LVD S | 2069 |
| Qy | 2356 | VHSE----GDYHRQTPGAWERDPSSTGSTOFYNNPLTMRLMISTPTPTACAPS AVNOA    | 2411 |







Db 863 SSMITQSPVPSQSPVSSSTTPSPGTLTSTSPSPSQSTTTIGTQGSTSPGISTTS 922  
Qy 1263 SPVASLEGICRALPRGSPHSDLKERTVLGSIHQGTTPRATTESFEDGLKPKQIKRES 1322  
Db 923 EEMTS--QG---STQTPGSTGTVPQSTVSDSTSSGS----- 955  
Qy 1323 PPIRFAEAGIYKRGPDYDITIKEMGRSHEIIPRODILITQSRKTPVQVSTRPIEGSI 1382  
Db 956 -----TWTVGSPGSSSPITPSTQNTN--PSTSSGSSMSTQTPQSSQSTP--VESST 1004  
Qy 1383 SOGTPIKDNNSGQAIKHNKSLITGSKLSRGNPPLEIPENIKVVERGKYEDV--KAG 1441  
Db 1005 SGAT-----SSSGSGT--TLTISPSPPSTIGSSOGSTSPVVSTISQGSTETPGSTG 1057  
Qy 1442 ETVRSRHTSVVSSGSPVLSRSTLHEAPKQALPGIYDDTSARRTPVSYONTMRSGSPMNR 1501  
Db 1058 STV--TKPSTVSGSAGSTATMGSTFASSTSG-----SSTSPNPSQST---SP---S 1103  
Qy 1502 TSDVTIPNKNSTNHERKSTLTPTQTORESIPAKSPVPGVDPVVSHPDPPHHRGSTAGEYVW 1561  
Db 1104 TSGATSSPGSGT--TLTISPSPPSTIGSSOGSTSPVVSTTSGDMTSQGST----- 1155  
Qy 1562 SHLPQLDPAMPFHRALPDAAAYLFQRLSPTPCYPQYQYLIYANETROIILNDIYTSQ 1621  
Db 1156 -QIPGSTGTV-----TOPSTGSGSTSGEITSQGSTQTPRSSLSSTP 1198  
Qy 1622 OMQVNLRPDVARGLSPREQPLGLYPAT--RGIIDLTNMPPITLVPHPGCTSTPPMDRITY 1680  
Db 1199 AISTSTQSVSTN--SPGST---VTQPSIVRG---STSGSVITVGTSGSTSGSSATS 1251  
Qy 1681 IPGTQITPPRPYNASMPGHPHTLAAASAEERERERERERERERERERERERERER 1740  
Db 1252 LSSSS-----PVPSTQSPNPST-----SGSSTPTNP 1279  
Qy 1741 G-----SEQPRGSGHYRSPSP--SVRTOETMLOQRPSVFOGTNGTSVITPLD--- 1788  
Db 1280 SOSTSPVVSTTTGEMTSHGTSQTPTIGSTVTQ-----PSTVSGSSSGSTVTIGSSE 1332  
Qy 1789 -----PTAQLRIMPLPAGGSPISQGLPASRY-----NTAAD-----ALAAVDAAS 1830  
Db 1333 ASTGSSFKTSPSSISPVPTSP-----IPSTTFASSTSGSTISDVSVSTISLAPLSS 1387  
Qy 1831 APQ-----MDVSKTKESKHAARLEENLRSAVSEQQOLEQKTLKVRKSVQCLYTS 1884  
Db 1388 LPSTVPSSTQSFSTSESSKASSPVSQTSSTPTNPTGSTESLTSSTISGTSQHTT 1447  
Qy 1885 -SAPSGRPQPHSSVYSEAGKDGKPPKRYEELRGTITTAANFIDVIITQIATAS 1943  
Db 1448 MSKASSGSTSPSTN---SQTG-----STVTMGSSSTSGVSTSSAS 1484  
Qy 1944 DKDARERGQSSDSSLSHRYETPPSDAIEVISPASSAPPOE--KLOTQOPEVVKANQ 2001  
Db 1485 STQPMNSTSQSSAGSTVAS-----STASPAASSTAPSTGTMSTSSSTSGTVGSTI 1534  
Qy 2002 AENDPTROYE-----GPLHRYRQOESPSPQOQLPPSPQAEQMGQ----- 2041  
Db 1535 SESSSTASASQSTGTVTWMGSSSTSGVSTSSASTQPMNSTSQSSAGSTVASSTAGLVS 1594  
Qy 2042 ---VPRTHRLI-----TLADHICQIITQDFARNQ-----VSSQTPOQ 2076  
Db 1595 TSTVPSSTGTMGSSSTSGVSTISSESTASASSQSTGTVTWMGSSSTSGVSTSSASTQ 1654  
Qy 2077 PTSTFQ-----NSPSALVST---PVRTKTSNRYSPESQAOQSVHHQRPGRSVRSPENL 2124  
Db 1655 QMSTSQSSAGSTVASSTGLVSTSTVPSSTGTMGSSSTSGVSTISSESTASASSQSTG 1714  
Qy 2125 VDKSRGSRGKSPERSHVSSEPEYIPSPQVPPVHHEKQDSLILLSQGAEPAEQRNDARS 2184  
Db 1715 STVTMGSSSTSGVSTSSASS-----GQPMNSTSQSSAGSTVVSST--ASPAASSTAPSS 1767  
Qy 2185 PGSTSYLPS-----FTFKLENTSPMVKSKQEIFRKLNSGGGSDMAAOPGTEINLPAV 2241  
Db 1768 TGTWSSSTSGVSTMSQSSAASTTSHTGTVTILGSSSTSSNMSTSQ--GSSVSGSTVAS 1826

Qy 2242 TTSGSVSRGHSFADPASNGLGLEDIIRKALMGSFDDKVEDHGVVMSQPMGVYP--GTANTS 2300  
Db 1827 STAGLVST-----STVPSSTGTMGSSSTGTSSTTASASSQSTGTYTMGSSSTS 1881  
Qy 2301 VYTSETRREGDPSPHSGGVCKPKLISKNSRKRKSKSPIGO--GYLGTFRPSSVSVHSE 2359  
Db 1882 GYSTSASSTQPMSTQSSAGSTVASSTAGLVSTVSTVPSSTGTMGSSSTGTVSTISE 1941  
Qy 2360 GDYHRQTPCWAWEDRPSSTGST---QFPYNPLTMRMLSSPTPTPIACAPSAVNOAAPH 2414  
Db 1942 SSTAASST-----SSQTGSTVTIGSTGNTSPRSLSQLITTP---SPS---OSTES 1987  
Qy 2415 QONRIWEREPAP-----LLSAQYETLS 2436  
Db 1988 TOTSLPSSPSPSTHVSSEGTTMS 2013

RESULT 34  
B35098  
MHC class III histocompatibility antigen HLA-B-associated protein 2 [imported] - huma  
C:Species: Homo sapiens (man)  
C:Date: 10-Aug-1990 #sequence\_revision 06-Nov-1992 #text\_change 15-Sep-2000  
C:Accession: B35098  
R:Banerji, J.; Sands, J.; Strominger, J.L.; Spies, T.  
Proc. Natl. Acad. Sci. U.S.A. 87, 2374-2378, 1990  
A:Title: A gene pair from the human major histocompatibility complex encodes large pr  
A:Reference number: A35098; MUID:90192810  
A:Accession: B35098  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2142 <BAN>  
A:Cross-references: GB:M33509; NID:gl79338; PIDN:AAA35585.1; PID:gl79339; GB:M31293  
A:Note: the authors translated the codon AGT for residue 97 as Gly

Query Match 2.8%; Score 353; DB 2; Length 2142;  
Best Local Similarity 20.4%; Pred. No. 3.4e-06;  
Matches 439; Conservative 224; Mismatches 833; Indels 656; Gaps 104;

Qy 503 QOIARPS--QBEKVBEKEEDKAEKTEKKEEKKDEEKD-----EKEDSKENTWEKDKI 554  
Db 299 EPVGRPSLKEDNL-----KEFDQLQDENDDGWAGAHEEVDYTEKLKFSDEE 345  
Qy 555 DCTAETE---EREQATPRGRKTANSQGRKRGRITRS--MTNEAAAAASAAAAATE--- 605  
Db 346 DGRDSEGAEGHRDSQSASGEERPEADGKAGNSPNEPPTPKTANAETSRPPTEPGP 405  
Qy 606 -EPPPLPPP-----PPEPSTPEVETSRWTEEMEVAKKGLVEHGRNMAATAK 652  
Db 406 PAPKPPLPGDYDPRGGPPCKPPAPEDEDEAWRQRKSSSEISL--AVERAR----- 456  
Qy 653 MYGTSKSAQCKNFYNYKRRHNLNLQHKOKTKSRKPREERDVSQCESVASTVSAQDEE 712  
Db 457 -----RRREERRRQEEERRAACAKL-----KRLDE 483  
Qy 713 DIEASNEENPEDSEVAVKPSDESPENATSR---GNTEPAVELEPTTETAPSTSPSLAV 769  
Db 484 KEGADPKRLKAEPAAPAA--PSTPAPPVAVPKELPAPPAPPASAPPTETEPE--EPAQAP 541  
Qy 770 P--STKPAEDESVEQVNV-----DSISAEPTAQMDVDVQOQHSAE--EGSVCDPPPAT--- 817  
Db 542 PAQSTPTPGVAAAPTLVSGGSGSTSTSSGSGFEASPVLPQPSKEGPEPEEVPPTTPV 601  
Qy 818 -----KADSDVDEVRVPHENHASKVEGDNTKRDLDRASEKVEPRDLDLVAAQIINAQRE 872  
Db 602 PKVEPKGIGIGTPRQP-----SQGLGPKYQKSL---PPRFQROOQEQLLKQO---QOHO 651  
Qy 873 PQSDNDSSATCSADEVDGEPERQRMFMDSKPSLLNPTGSLTVSSPL--KPNPLDLPLQLO 931  
Db 652 WOHQOQGSAPPTPPVPPQPVTLGAVPAPQAPP---PPPKALYPGALGRPPMPMFMFD 708  
Qy 932 HRAAVIPPMVSCPCNIPITGTPVSGYALYQRHKAHESALLEEORQ-----RQEQIDLE 986



|    |      |                   |                   |                 |                               |                      |               |     |
|----|------|-------------------|-------------------|-----------------|-------------------------------|----------------------|---------------|-----|
| QY | 365  | RGAGLSATIARSEHIEI | SEI               | IDGLSEQBNNEKQMR | -QLSVIPPMFMFAEQRRVKFINNM-     | 422                  |               |     |
| Db | 150  | -----AERIRGEDEKS  | -----MLAKEHADKIRL | GVAKIPRLLT      | ESERKMDEBEPVGS                | 198                  |               |     |
| QY | 423  | LMEDPMKYK         | ----              | DRQFMVWTDHEKEIE | KDFIOHPKNFGLIASYLERKSVPDCVLXY | 478                  |               |     |
| Db | 199  | ILKDMKKEHQSVLDR   | --LWESPEERSL      | FKSQAODHVKIFLGH | LTEFFVDKTASDLVLVY             | 256                  |               |     |
| QY | 479  | YLTKKNENYKALVRN   | YKGRGRNQI         | TARPSOEKVEEKEE  | DKAEKTEKKEE                   | 530                  |               |     |
| Db | 257  | YMNKTEDYKDFPK     | --KRVTKYKCAP      | FSVEELAYFRMMP   | PLDRSSFPKNSLMCYFC             | 314                  |               |     |
| QY | 531  | -----BKDEEKEDEK   | ESKENTKEKDI       | -DGTAEETEERQ    | APRGRKXTANSOGRK               | 582                  |               |     |
| Db | 315  | RTVANGIDLNGTFM    | PEKAEYEIFAICPD    | EDRVVCGSREAAK   | LKDNRCFGNCSNOKKRA             | 374                  |               |     |
| QY | 583  | GRITRSMTNEAAAA    | AAAAATEEPPPL      | PPPEP1STEPET    | SRWTEEMEVEAKKGLVE             | 642                  |               |     |
| Db | 375  | NRVNRNI           | -----PLD          | ADPVRTRAFIMDK   | LGSTR                         | 404                  |               |     |
| QY | 643  | HGRNWAIAKMGV      | TKSEAOCKNFY       | -----FNTKR      | RHNDLNLQOHQ                   | 686                  |               |     |
| Db | 405  | -----VAVFCT       | PCKNALTRWINDVN    | KNKEETIMAE      | LLNYSQGVGTDDKTKLV             | 453                  |               |     |
| QY | 687  | -----SKPREERDV    | SOCESVASTVS       | QAQEDDETEA      | NEENPEDSEV                    | 728                  |               |     |
| Db | 454  | TLINSSPTLDWVS     | ISEGMNRPRNECK     | MOYDAMNGVKT     | QPMIEVDEEDNGOEGGDALV          | 513                  |               |     |
| QY | 729  | EAVRPSDS          | -----PENAT        | SRGNTPEAVELE    | PTTETATSTPSLAV                | STKPAEDSVET          | 782           |     |
| Db | 514  | NPTTSSAAARRS      | GLARNAKKPVTR      | PRAPSAGRR       | TGGAVTRAQAVP                  | --KPED               | 566           |     |
| QY | 783  | QVNDISIAETAQ      | MDVQDOEHS         | AEESG           | -----VCDPP                    | PATKADSDVDVVRV       | PENHASKVEGD   | 839 |
| Db | 567  | -----LGEIEDE      | IEDNEDEASRGS      | RKDSKAPSDR      | SGSPADMEGSD                   | PEGDQDQDQDQ          | 620           |     |
| QY | 840  | NTKERDLDRASE      | KVEPRDEDLV        | VAQIINAORPE     | QSDNDSATCS                    | DADEVDGDEPERQRMF     | 899           |     |
| Db | 621  | QDQDQDVEE         | EEVIVRIDIS        | PKVTKLL         | -----SPKIL                    | SGGHRKDPDP           | 663           |     |
| QY | 900  | PMDSKPILLNPT      | GSILVSSPLK        | NPLDLPOLOHRAA   | -----VITPMV                   | SCTPCNIPITGPV        | 954           |     |
| Db | 664  | PRIQKS            | -----TSQ          | PPPEPMDTKENES   | DGDEENDILEID                  | VDPPAKRPTPTSS        | 715           |     |
| QY | 955  | SGYALYQRHI        | -----KAMHES       | ALLIEQRQEQ      | IDLECRSTSC                    | PGTSKSPNREWEVLQPA    | 1011          |     |
| Db | 716  | SSHILIGSSVGS      | RELGRLVQ          | QOQOQOQOQ       | QAQSAAPPVTV                   | STAAAAAERLVNATSPS    | 775           |     |
| QY | 1011 | PHQLITNLPEGV      | RLPT--TRTR        | PRPPPLIPSK      | KTIVASEKPS                    | FTMGSSISQGTGPGYILTSH | 1061          |     |
| Db | 776  | PSVASQHL          | -----VPTET        | STSVPPVTVPP     | -----AIQ                      | PVVVISAQAQFT         | --QOQO        | 823 |
| QY | 1070 | NOASVTOBTP        | -----KPSVGSIS     | -----LG         | PROQESAKS                     | SATLTP               | 1101          |     |
| Db | 824  | SPALIAQBIPO       | OLLIPQVR          | SVTPAQILPT      | PVRPTAA                       | STPSMDQFLGLFKQOQ     | 874           |     |
| QY | 1105 | IKQEEFPSRSON      | SOPE-----GLLV     | --RAQHEGV       | VRGTAGATIQ                    | -EGSITRGTPT          | 1141          |     |
| Db | 875  | -QOQOQOQ          | PQOSNLMOQL        | GNINPQFLALL     | LQOQOQOQOQ                    | QAQVQAQATQ           | SGTSTSGTPTFOA | 933 |
| QY | 1150 | -----TSKIS        | VESIPSRG          | SITOGT          | PALPQ                         | 1174                 |               |     |
| Db | 934  | QORFEALQK         | LFSSPEMLG         | TLLN            | AKYQFPQ                       | 963                  |               |     |

RESULT 37  
T34513  
hypothetical protein ZK783.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34513  
R:Favella, A.; Vaudin, M.  
submitted to the EMBL Data Library, August 1994

[illegible]

Db 2431 TASEPTKSTMSPLSTTSNVLSESTPESSK--SPVSSSTEGISVTSTEFVKVPE 2488



```
QY 997 -----SKSPNREWEVLQAPHOLITN-----LPEGVRLPT-----TRP 1029
Db 1664 IRQMERVPWFVSKSRWE-----PNKVSNNSGMLPNAVLPPSLDHNAYQWQERESSH 1718
QY 1030 TRPPP-----PLIPSKTTTAVASEKPFIMGSGISQGTPTGTLTSHNQASVQBTPTKPSVG-- 1084
Db 1719 TEQPLMKKIIIPAKPKPGEPDS-----PTPLHPPTPTLSTRKEDSPELNPPPGIDDN 1775
QY 1085 -SISLGLPQROESAKSA-TLPIYIKQEFSP-----RSONSQPGLLVRQAHEGVWVG-- 1134
Db 1776 RQCALCLMIGDSDANDAGRLIYIQONEWTHVNCALWGAEEVDEDDGSLKNVHMAVIRGKQ 1835
QY 1135 -TAGAIOGSGTTRTPPTSKISVESIPSLRGSITQGTALPQTGIPTEA---LVKGSISRM 1190
Db 1836 LRCBFCQKPGATVCCCLT---SCTSNYHFMCSRANKCVFLDDKKYICQHRDLIRKEV--- 1890
QY 1191 PIEDSSPEKGREEAASKGHVIEGKSGHILSYDNINKNAREGTRSPRTAHEISLKRSYESV 1250
Db 1891 -----VPENG-FEVRFRVDFEG-----ISLRKF--L 1916
QY 1251 EGNIKQMSMRSPVSPAPLEGLI-----C--RALPRGSPHS-----DUKERTVLSGSI 1296
Db 1917 NGLPEPNIHMGSMWIDCLGILNDLSDCEDKLPFIQYQCSRVYTWSTTDARKRCVYTCKI 1976
QY 1297 MOGTP-----RATTE-----SPEDGLKYPKQIKRE--SPPIRAFEGAITKG 1335
Db 1977 MECRPPVVEPDINTVEHDDNRTTAHSPSSFDASCDSQTAAILPSPDRPHSQTSQ 2036
QY 1336 KPYDGIITIKEMGRSIIHEIP-----QDILTOES---RKTPEVQOSTRIPIBGSISQGTPI 1388
Db 2037 SCYVHV-----ISKVPRIPTSPYQTSRQPCGPLPSAGSPT-PTTHEIVTVGDPL 2086
QY 1389 KFDNNSGOSAI---KHNVKSLITGPKSLRGMPPLEI-----VPE----- 1425
Db 2087 L---SSGLRSIGSRHRSTSSLSLRSKL-RIMSPVRTGSAYSRSVSPSLGATDPEA 2142
QY 1426 NIKVVERGKYEDVAKG-----ETVRSRHTS-VWSSGSPVLR--STLHEAPKAQ 1470
Db 2143 SAKASDRGGLLSSANLGHSAAPPSSSQRTVGGSKTSHLDGSSPSEVKRCSALDLVPKGS 2202
QY 1471 LSPGIYDDT-SARPTPVSYQNTMSRG-----SPMNRNTSDVTIPPNNKTNHERKSTLTPTQ 1525
Db 2203 LVGKKNRTSSSKTGDGSAHSTAYPGIPKLTQVHNATPGELNISKIGSKIAEPSTVPFSS 2262
QY 1526 RESI-----PAKS--PVPGYD-----PVSHSFDPH-HRGST 1555
Db 2263 KDTVSYFQHLRGORSRDQHMDSQSVKPSPNEDGEIKTLKPGMGRHPSILHEHGSS 2322
QY 1556 AGEVYWSHLPTQLDPAMPFHRA---LDPAAAAYLFRQLSP-----TPGYPSQYQLYA 1605
Db 2323 SRDRRQKGGKSKETCKEKHSSKSYLPGQVTTGEGNLAKEPADEVLTGFLGQ---RP 2379
QY 1606 MENTROTILNDYI-----TSQOMQVNLRPDVARGLSPREOPLGL----- 1644
Db 2380 CNNVSEKIGDKVLPISGVKPGQSTQVEGSKELQ-----APRCKSVKVTPLKMEGENQ 2433
QY 1645 -----PYPATRGIIDLTNMPPILVPHPGGT----- 1670
Db 2434 SKNTQKSGGSPAHIESVCPAEPVSASRS-----PGAGPGV-QSPNNTLSQDPQSNYQ 2488
QY 1671 STPPMDRITYIPGTQITFPPRPYNASMSGHPHTHAAAAS----- 1711
Db 2489 NLPEQDRNLMPD-----GPKPQEDGSKFRYPRRSARARNMFFGLTPLYGVRSYGEED 2543
QY 1712 -----AER 1753
Db 2544 IPFYNSNGTKRGRGRSGAGQVGDGADLSTDEDDLYNYFNTRTVISSGGEE--RLASHNL 2601
QY 1754 VRSPSPSVRTQETMLQORPSPVFOGT-NGTSVITPLDPTAOLRIMPLPAGGSPISQGLPAS 1812
Db 2602 FREEE---QCDLPKISQDGVDDGTSTSVTSRKSQI---PKRNGKNGTENLKID 2655
```

```
QY 1813 RYNTAADALAAALVDAAA--SAPQMD-----VSKTKESKHEA-----ARLEENLRSRSAVS 1861
Db 2656 RPEDAGEKEHVIKSAVGKHKNEPKLDNCHSVSRVKAQODSLEAOLSSLESSRRVHTSTPS 2715
QY 1862 EQOOLEKOTLVEK-----RSVQCLYTSSAPPSG-----KQPHSSVYVSEAGKDKGP 1909
Db 2716 DKNLLDTYNAELLKSDSDNNNSDDC---GNILPSDIDMFVLKNTPSMQAL-----GE 2764
QY 1910 PPKSRYEEELRTRGKTTITAAFNIDVILITQIASDKARERGSSOSSSSSSSLSHRRETTP 1969
Db 2765 SPSS--SEELLTGEGLDLSNRKEDIGLFEVFSQQLPATEPVDVSSVSSISAEBEQFELP 2823
QY 1970 SDA---TEVISPASSPAPQ-----EKLOTYQPEVVKANOAEENDPTROYEGPL 2014
Db 2824 LELPSDLVLTIRSPTPVPSQNPRLAVISDSGERVITTEKSVAS--SEGDPALLSPG-- 2879
QY 2015 HHYRPOQESPPQQLPPSQOAGMGQVPRTHRLITLADHI---CQII-----TDQFA 2064
Db 2880 -----VDPAPEGHMTDPHFIOGHMD-----ADHISPPCGSGVEQGHGNSQDLT 2922
QY 2065 RNQ-----VSSQTPOQ-----PPTSTFQNSPS---ALVSTPVRTKTSNR----- 2100
Db 2923 RNSGTPGLQVPVSPTPVQNKYVPSTDSFGPSQISNAAVQTTTPHLKPATEKLIIVNQ 2982
QY 2101 -----YSPESQAQSVHHQ-----RPGSRVSP 2121
Db 2983 NMQPLYVLQTLPNGVTQKIQLTSPVSTPSPVMTNTSVLPGMSGGLTTLTGLNPSLPPSP 3042
QY 2122 ENLVKSRGRSPGKSPERSHVSSEYEPIS--PPQVPVWHEKQDLSLLLSQGAEP----- 2175
Db 3043 SLFPPASKGLL--SVPHQHHLHSFPAAQSSFPN---ISSPPSGLLIGVQPPDPQLLG 3097
QY 2176 --AEORNDARSPGISVILPSEFFKLENTSPMKSKKEIEPRKLNSSGGSDMAAAQ--- 2230
Db 3098 SEANQRDLTTTVA---TPSSGLKKRPIRLHTRKNKKLAPSSAPSANIAPSDVVVSNMTLI 3154
QY 2231 --PQTEFNLPNAVITSGSVSRGSHFADPASNLGLEDIIRKALMGSDDKVEDHGVVMSQ 2288
Db 3155 NFPTPSQLSNHPSLLDLGSL-----NPSHRIVPNLIKRKSGIM--YFEQAPLPPQ 3204
QY 2289 PMGVVPGT-ANTSVVTGSTRREBDDPSHGGVCKPKLISKNSR--KSKSPITPG----- 2341
Db 3205 SVGTAATAAGSSITISQDTSHTLTSGLPVSAALSSGVNLVNVSMQTAAPTSTSVPGHVL 3264
QY 2342 --QGYLTERPSSVSVSHSEGDYHRTQPGNAWEDRPS--TGSQTFPYNPLTMRMLSTP 2397
Db 3265 ANQRLLETPDIGSLSHLILKASH--QSLG--IQDQVALPSSGGMFFQ-----LGTS 3312
QY 2398 PTPIACAPSAVN 2409
Db 3313 QTPSAAAMTAAS 3324
RESULT 39
S33124
tpr protein - human
N:Alternate names: kinase-related transforming protein (tpr-met); protein with promot
C:Species: Homo sapiens (man)
C>Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 17-Mar-2000
C:Accession: S33124; S23740; S00928; G01185
R:Mitchell, P.J.; Cooper, C.S.
OncoGene 7, 2329-2333, 1992
A:Title: The human tpr gene encodes a protein of 2094 amino acids that has extensive
A:Reference number: S33124; MUID:93064711
A:Accession: S33124
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-2094 <M1>
A:Cross-references: EMBL:X66397; NID:g633225
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
R:Mitchell, P.J.; Cooper, C.S.
OncoGene 7, 383-388, 1992
A:Title: Nucleotide sequence analysis of human tpr cDNA clones.
```

[illegible]





QY 1479 TSARTPVSYQNTMSRGSPMMNRTSDVT---IPPNKSTNHKST-LPTQRESTPAKSP 1534  
Db 1159 SLP-----EGATSPGPRREAPPQVCGWSPPAKSLAPKKPTGTLPLPPSKLEKEL- 1210  
QY 1535 VPGVDVYSHSPFDPHHRGSTAGEVYWHLTOLDPAMPFHALDPAALYLFQRLSPT 1594  
Db 1211 IFG-----PLSPVARGSGNG---GNGVME-DGERPRRRHGRA-----QQQ----- 1248  
QY 1595 PGYPSQYQYAMENOTRQTLNDYITSOQMVLNRPDVAGLSRPRQPLGLPYAPTRGIID 1654  
Db 1249 -DKPPFRRLKQE-----RENAARG-SEGKPSLTLPASA----- 1280  
QY 1655 LTNMPTTLVPHPGTSTPPMDRITYIGTQITPPRPNYNSASMP-----GHPTH 1705  
Db 1281 -----PG-----PEAL-----TTVTAPAPRAAAKSPDLNSQNSDQANBEW 1318  
QY 1706 LAAASAEREREREKE-----LPAASYNTAADALAAALVDAASAPQMDVSTKESKHEAA 1847  
Db 1319 ETASESDFTSERRGDKKAPPPVLLTPKAVTGGGGGAVPGISAMSRGDLISQRAKDL 1378  
QY 1738 LRPGEQPGRFGSHGYVRSPSPSVRTQETMLQORPSVFQGTNGTSVITPLDPTAQLRIMP 1797  
Db 1379 KRFSFQ--RPGMERQRRPGGKA-----GSSGSS-----SG 1410  
QY 1798 LPAGGPSISQ-----LPASRYNTAADALAAALVDAASAPQMDVSTKESKHEAA 1847  
Db 1411 GAGGPGGRTGGRGDKRSWSPKNSRPPBE-----RPPGLPLPPPPSSSAVF 1460  
QY 1848 RLEENLRSAVSOOLEQKTLEVERSVQCLYTSSAFPSGKPOPHSSVYVSEAGDK 1907  
Db 1461 RLDQVHNSPAGI--QQALAQ--LSSRGSVT---APGGHPRHKGPG-----POAPQGP 1507  
QY 1908 GPPKRSYEEELTRGKTTITAAANFIDVITROIASDKADARGSQSSDSSSLSSHYE 1967  
Db 1508 SPRPTRYEPQVRNSG-----LSSDFHFEPEGVPVRGVG----- 1542  
QY 1968 TPSDAIEVSPASSPAPQOEKQ--TYOPEVYKANOENPTROYEGPLHHYRPOQESPSP 2026  
Db 1543 TPRDSAGV-----SPFPKRRRPPKPELL---QESLPPPHSSGFL---GSKPEGPGP 1591  
QY 2027 QOOLPPSSQAEQMGQVPRTHLLTLADHICQIITQDFARN--QVSSQTPOQPPSTFQ-- 2082  
Db 1592 QAE-----SDTQTEALTPHWNRLHATSKRYRPSMPEPMEPLSPFEDV 1638  
QY 2083 -----NSPSALYSTPVRTKTSNRYSPESQAQSVHH---ORPGSRVSPENLV 2125  
Db 1639 AGTEMSQSDSGDLSDGQVSSGP-----CSQSSPDGGLKGAEGPPKRGGS-SPLNAV 1693  
QY 2126 DKSGRSGKSPERSHVSSEPEYIPSPQVPVHVEKODSLLLLSQR-----GAEPABQR 2179  
Db 1694 -----PCEGPP-----GSEP--PRRPPAP--HDGDRKELPREQPLPPPIGTERTSQR 1738  
QY 2180 NDARSPGISVLPSTFFTKLENTSPVK-----SKQEIFRKLNSGGGSDMAA-----AQP 2231  
Db 1739 DRATEPGPIR--PS-----HRGPPVQFGTSKQSDSLVVGDSLKAEKELTASVTEAIP 1791  
QY 2232 GTEIFN-LPAVTTSGVSRR-----GHSFADPASNLGLIEDIIRKALMGSGFDDKVEDHGVVM 2286  
Db 1792 VSRDWELLPSAAAEQSKNLDGSHCVPEPSS--GQRLYPEVFGS-----AGPSS 1842  
QY 2287 SOPMGVPGTANTSVTSGETRRREGDPSH 2317  
Db 1843 SQ-----ISGGSHGLSITSKQWRLRPGTPSLH 1869

## RESULT 41

S13507  
microtubule-associated protein MAP2 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 13-Aug-1999  
C:Accession: S13507  
R:Marchal, D.; Delapierre, D.; Dresse, A.  
Arch. Int. Physiol. Biochim. 96, 231-236, 1988

A:Title: Cloning and partial sequencing of a new rat brain specific cDNA.  
A:Reference number: S13507; MUID:89334524  
A:Accession: S13507  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-1825 <MAR>  
A:Cross-references: EMBL:X54100; NID:g56624; PID:CAA38034.1; PID:g56625  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1990  
C:Superfamily: microtubule-associated protein MAP2b; MAP2/tau repeat homology  
C:Keywords: microtubule binding; tandem repeat  
F:1667-1697/Domain: MAP2/tau repeat homology <MT1>  
F:1698-1728/Domain: MAP2/tau repeat homology <MT2>  
F:1729-1760/Domain: MAP2/tau repeat homology <MT3>

## Query Match 2.7%; Score 338; DB 2; Length 1825;

Best Local Similarity 20.1%; Pred. No. 1.1e-05;  
Matches 359; Conservative 232; Mismatches 587; Indels 606; Gaps 94;

QY 1008 QPAPHQLITNLPEGVRLPTTRTPRPPLIPSSKTTTVAASEKPSFIMGSGISQGTPTGLT 1067  
Db 119 QPAALPLAAE--ETVNL-----PSPPPSPASEQTAALAEAS----- 153  
QY 1068 SHNQASYTQETPKPSVGSISLGLPRQESAKSATLPYIKQE-EFSPRSONSOPEGLLVRA 1126  
Db 154 -----KMEFPQOKLPSSFABPLDKETEETEFKQSKPGED-----F 188  
QY 1127 QHEGVV--RGTAGAIQ-----EGSITRGTPTKSIVESIPSLRGSITQCTPALPQTG 1176  
Db 189 EHAALVPODTSKTPQDKKDPQDMEGEKSPASFPATQGTGNLEDIK-QITEPSITVPSIG 247  
QY 1177 IPTALV---KGSIRMPIDESSPEKREEAASAGHVI-----YEKG--- 1215  
Db 248 LSAELAPKQDKQDWIEMPVESKKDEWGLAAPISGGLTPMKREKDVLEIDIPWEGKQFDS 307  
QY 1216 -----SGHI--LSYDNINKNAR--EGTRS-----PRTAHEISLKRYSYEV 1251  
Db 308 PMPSPFHSGSFTLPDITVKDERVTEGSPFAPVFFQSDDKMSLQDTSGSATSKESKQBE 367  
QY 1252 GNIKQMSRSPVSAPEGLICRALPRGSHSLDKERTVLSGSIWGTTPRATTESFEG 1311  
Db 368 POKKADKADVDPVSEATTVL-----GDVHSPAVEGFV--GENISGEKGTID-- 413  
QY 1312 LKYPQIKRESPPIRAFEGAITKGPYDITIKEMGR-SIHEIRQDILITQESKRTPE- 1369  
Db 414 -----QEKETSTPSVQETLITETEPQ---TKLETSKVSEETVAKE---EESLKLDD 462  
QY 1370 ---VVQSTRPIEGSISQGTPIKFDNNSGQ-----SAIKH-----NVKSLITGPSKLSRM 1417  
Db 463 KAGVIQTS---TEHSFSK-----EDQKGEQTIKALQKDSFPISLEQAVTDAAMATKL 513  
QY 1418 PPLEIVPENI--KVVERGKYEDVKAGETVRSRHTSVSSGSPVSLRSTLHEAPKAQLSPGI 1475  
Db 514 EKVTSPEAVSEKREIOGLFEEDIADK-----SKLEGAGSATVAEV--EMPYEDKSGM 565  
QY 1476 --YDTSARRTPVS-----YQNTMSRGSPMMNRTSDVTIPPNKSTNHKSKTTLPT 1524  
Db 566 SKYETSALKEDVTRSTGLGSDYELSDSRGN--AQESLDVSPKNQ---QDEKELLAKA 620  
QY 1525 QRESIPAKSPVGVDPVYVSHSPFDPHHRG-STAGEVYVSHLPTQL--DPAMPFHR--ALD 1579  
Db 621 SQSPPPA-----HEAGYSTLAQSYTSDHPSELPEEPSPSSPOERMFTID 662  
QY 1580 PAAAYLFQRLSPTPGYPYQYQLYAMENOTRQTLNDY-----ITSQQQVNL----R 1628  
Db 663 PKV-----YGEKRDLSHKNKDDTLRSRLGLGGRSAIEGRSMISINLPMCL 708  
QY 1629 PDVARG-----LSPREQPLGLPYPATRIIDLTN--MPETILVPHPGTSTPMDRI 1678  
Db 709 DSIALGFNFGRHDLSPASDI---LNTVSGTMDGDDYILPPT-----TPAVEKI 755  
QY 1679 TYIP-----GTQIT-----FPPRPY--NSASMSPGHPTHLAAA-- 1709



[illegible]

RESULT 43

RESOL  
T45025

145025  
mucin MUC5B, tracheobronchial [imported] - human (fragment)

C;Species: Homo sapiens (man)

C;species: homo sapiens (man)  
C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000

C;Accession: T45025

C/Accession: 145025  
R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.

K. Bessy, V. L., Guyonnet Duperrat, V., Fournet, N., Aubert, J. F., Baudin, A.  
J. Biol. Chem. 272, 3168-3178, 1997

**A;Title:** Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter

A:Reference number: Z22899; MUID:97166151

A:Reference Number: T45025

A;Accession: T43023  
A;Status: preliminary: translated from GB/EMBL/DBJ

A: status: preliminary  
A: molecule type: DNA

A: Molecule type: DNA  
A: Residues: 1-3570 <DES>

A;Residues: I-3570 <DES>  
A:Cross-references: EMBL:272496: NTD:α1834502: PTD:CAA96577.1: PTD:α1834503

A:Cross-references: EMBL: Z/2496  
A:Experimental source: placenta

A; Experiment  
C; Genetics.

C;Genetics:  
A:Gene. MTIC5B

Query Match 2.7% Score 336 DB 2 Length 3570:

Query match 2.7%, score 330, DB  
Best Local Similarity 18.9%: Pred. No. 3e-05

Best Local Similarity 18.5%, Fied. NO: 3E-05;  
Matches 410: Conservative 228: Mismatches 830: Indels 696: Gaps 91:

OV 567 ATPGRKTSNGRRKGRITRSMTNEAAAAAATAEEPPPLPPPPPEPISTEPVETS 626

QY 307 AI PRGRATANSQGRKRGRTT RSMINEAAAASAAAAATEEPFPLPFPFEPISTEFVEIS 020

DD 399 AISKARFPGIASTASKEPLTISLAPTILSELSTSQAEISIFRIETIMSPFLINITSQGIT 430

OV 627 R-----WTE-----EEMEV-----AKKGLVEHGRNWAAIA-KM 653

D6 459 RCOPKCEWTEWFDVDFPTSGVASGDMETFEENTRAAGGKMCWAPKSTIECRAENYPEVSDO 518

DB 459 RCQPKCEWTEWF DVDFPTSGVASGDMETFFENIRAAAGGRMCWAPKSIECRAENYPEVPSIDQ 518

QY 654 VG-----TKSEAUCK-----NFYFNFKRR-----HNLDNLLQVHKQKTSRRKPR 091

519 VGOVI.TCSCI.ETGL.TCKNEFOTGRENMCENYNVRVI.CCDDVSHCPSTI.ATSSSTATPSSSTPG 578



A:Residues: 1-5262 <PRA>  
A:Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2358285  
C:Genetics:  
A:Gene: ALR  
A:Map position: 12  
C:Superfamily: human ALR protein  
C:Keywords: alternative splicing

Query Match 2.64; Score 335; DB 2; Length 5262;  
Best Local Similarity 19.48; Pred. No. 5.3e-05;  
Matches 466; Conservative 267; Mismatches 882; Indels 786; Gaps 113;  
QY 566 QATPRGRKANSQGRKGRITRSMTEAAAAASAA-----AAATEE-----ppp---p 610  
Db 393 QGQPKGGHVTNQKPEG---PLOCAPKLGKAGVQLPEQLAPLNEEMPLPPPEESP 448  
QY 611 LPPPEPEISTE-PVTSRWTEEMEVAKGLVEHGRNAAIAKMGVTKSEAKCNFYNY 669  
Db 449 LSPPEESPTSPPEASRLSPPEELPASLPE-----AL----- 483  
QY 670 KRRHNLNLLQHKQKTSRKREERDVQCESVASTVSAQEDIEDIASENEENPEDSEVE 729  
Db 484 -----HLSPLEESPLSPPEESPLS-----PPPESSPFPLEESP----- 519  
QY 730 AVKPSDESPEATNRCNTEPAVE---LEPTTETAPSTSPSLAVPSTKPAEDESVEQVND 786  
Db 520 -LSPEESPPSPALETPLSPPEASPLSPPEESPLSPPEELTSPPEASRLSPPEE 578  
QY 787 SISAEATQMDVQOQHSAE-----EGVCDPPPATKADSDVDVEVRVPENHASKY----- 836  
Db 579 SPMSPPEESPMSPPEASRLSPPEESPLSPPEESPLS-----PPPEASRLSPPEE 631  
QY 837 -----EGDNTERDLDRAS-----EKVEPRDEDLVAAQINAAQPE-----PQSD 876  
Db 632 DSPMSPPEESPMSPPEESRLSPPLVVSRLSPPEESPLSPALSPLEGELEYFPKAGD 691  
QY 877 NDSSATCSA---DEDVDGEPRQMFMDSKPSLINPT-GS-ILVSSLPKPNPLD---LP 928  
Db 692 SDPESPLAAPILETIPSPPEANCTDPEVPPMILPPSPGSPVGPASPILMEPLPPQCS 751  
QY 929 QLOHRAVIP---PWSCTPCNIPIGTVPVGYALYQRIHKAHESALLERQOEIDL 985  
Db 752 LLOH--SLVPQNSPPSQSPALPLSVP-SPLSPIGKVVGVSDAEALHEMETEKVS---EP 806  
QY 986 ECRSSTSCGTSKSPNREWEVLQAPHQL-----ITNLPE-----GYRLTPTTRTPP 1034  
Db 807 EC-PALEFSATSPSPMGDLSCFAPSAPALDDFSGLGEDTAPLDGIDAP---GSQPEP 862  
QY 1035 PLIPSSKTTVASEKPSFTMGSSIISQGTGTLYLTSNQAAS-----YTQETPKPSV----- 1083  
Db 863 GOTPGSLASELKGSPVLDDPEELAPVTMEVYPECKQTAGRGSPCEQEEPRAPVATPP 922  
QY 1084 ----GSISGLPQOESAKSATLYIKQEEFSPRSQNSQPEGLLVRAHQEGVWGTAAGI 1139  
Db 923 TLIKSDIVNEISNLQGDASASF-----GSEPLLGSPDEGGGSLMELGVSTDVSPAR 977  
QY 1140 QEGSI-----TRCTPT-SKIS-----VESISLR 1162  
Db 978 DEGSURLCTDSLPTDSSLCDAGTATSGGKAEGEKGRRSSPARSRITKQGRSSFPGR 1037  
QY 1163 -----GSITQGTPALPQTGIPTAEALVKGISIRMPTEDSPEKGREP-----AASKGHVI 1211  
Db 1038 RPRGAHGRGRARLAKSTASTIETLVADI-----DSSPSKEEEDDDTQNTVVL 1091  
QY 1212 Y-----EGSGHILSYDNIKNAREGTRSPRTAHEISLKRYSVES 1251  
Db 1092 FSNTDKFLVMDQMCVGSFGRGAEGHLLACSQCQCVCVNSKITKVMLLKGRVCE 1151  
QY 1252 GNIKQGHSMRESP-----VSAPLEG-----LICRALPRGSP- 1282  
Db 1152 CIVCEVCGQASDPSRLLLCDDCDISYHTYCLIDPPLAVPKGGMKRCVCSMCQCAASPG 1211

QY 1283 -HSD-----LKERTVLGSGSIMOCTPRATTESF--EDGLKY- 1314  
Db 1212 PHCEWQNSYTHCGPCASLVTPCICHAPVEEDLLILOCHCEHMMHAGCESLETEDVDVHA 1271  
QY 1315 -----PKQIKRESPPIRAFEGA-ITRGKPYVDGITTICE 1346  
Db 1272 PDEGDCVSCQYVVKVPVAVAPPVPMKVKPEPEQYFRFEGVWLTE---TGMALLRN 1327  
QY 1347 MGRS-IHE-----IPRQ-----DILTOESRK-----TPEVVQSTRPI--TEG 1380  
Db 1328 LTMSPHLKRRORRGLGUPGEAGLEGSEPSDALGDDKDDGLDTELLKGGEGVEHREC 1387  
QY 1381 SISQCTPIKIFNNNSQSAIKHNK-----SLITGPKS-----LS 1414  
Db 1388 EIKLEGPSPDVEPGKEETESKRRKRPYRPGICGFVWRQKSHTRTKGPAQAQAEVLS 1447  
QY 1415 RGMPLLEIVPENIK---VVERGKYE-DYKAGETVRSRHTSVSS--GPSVL-----RSTL 1463  
Db 1448 GDGQDEVIPADLPAGAVEQSLAEGDEKKQORRGRKRSKLEGMPAYLOEAPFGKELL 1507  
QY 1464 HEAPKAQLSPGIYDDTSARRTPVSQYQNTMS--RGSPPMNRSTD-VTIIPNKSTNHERKST 1520  
Db 1508 DLSRKALFAVGVRPSFGLGTPKAKGGGSKERKELPTSQKGGDDGPDIADEESRLEGKAD 1567  
QY 1521 LTPQRESIPAKSPVPGVDVVSHPFHHRGSTAGVYVWSHLPTOLDPAMPFHRALDP 1580  
Db 1568 -TPGEDGVKASVPVS-----DPEKPG-TPGE---GMLSSDLDR----- 1603  
QY 1581 AAAAYLFORQLSPTGPGYPSOYQLYAMEN-TROTILNDYITSOQMOVNLRPDVARGLSPRE 1639  
Db 1604 -----STEELPKMESKDLOQLFKDVLGSE-----RE 1629  
QY 1640 QPLG-----LPYPATRIIIDLTMPPYI-LVPHFGGTSTPPMDR-----ITY 1680  
Db 1630 QHLGCGTGLEGRTRPQLQRFLOGLPLGNLFSSPMDSYFCLQCSPLDFRERGGFSP 1689  
QY 1681 IPGTQITFPPRPY-NSASMSPGHPTLAAAAASAE-----RERERER-EKERERERIAAASS 1734  
Db 1690 EPG-----PDSPTWTCGGTITPTTPTTEGEGDGLSYNQSLQWKEDEELGQLTISP 1745  
QY 1735 DLYLR---PGSEQGRPGSHGYVRSPSPSVRTQETM----- 1767  
Db 1746 VLYANINPNLKQD-----YPMWSSRCQIMKLWRKVPAAADKAPYLQKADNRAA 1795  
QY 1768 -----LQORPSVFOGTNGTSTVITPLD-PTAQLRI-----MPLPAGSPISOGILPA 1811  
Db 1796 HRINKVOKAESQINKOTKVGDIARKTRPALHURIPQPGALGSPPPAAAATFIGSPT 1855  
QY 1812 SRYNTAADALAAALVDAASAAPQMDVSKTKESKHEA-ARLEENLRSRAAASEQQOLEQT 1870  
Db 1856 TPAGLSTSDAGDFLAPPGASVPGPD-----SPGELFLKLPQVPAQAPSQDPGLAPAYP 1909  
QY 1871 LEVERKSVOCLYTSAPPSGPK-QPHSSVYVSEAGKQK-----GPPPKSYEE----- 1918  
Db 1910 LEPRFTAPPTYPYPSPTGAPAPQPMPLGASSRPGAGQPGEFHTTTPGTPRHQPSDPPF 1969  
QY 1919 LRTGKTTITANEDIVITITQIASDKDARERGSSQSSSSLSLSSHRYETFSDALEVISP 1978  
Db 1970 LKPCPS-----LD-----NLAVSPESGVGGKASE--PLLSPPPPGESRKALEVKKE 2015  
QY 1979 ---ASSPA--PPQ-----EKLQTYQEVVKANQANDPTROYEGPLHHVRPQ 2020  
Db 2016 ELGASSPSYGPNNLGFVDSPPSGTHLGLGLELKTDPDVFKAPLTPRASQVEPQSGLGLRPQ 2075  
QY 2021 QESPSPOQLPPS-----SQAEGMGQVPRTHRLITLADHICQIITQDFARNQVS 2069  
Db 2076 E--PPAQAALAPSPPSHPDIFRPGSYTDPYAQPPLTPR----- 2111  
QY 2070 SQTQOQPTSTFQNSPALYSTVPTRTKTSNY-----SPESQ---AQSVHHRQPSGRVS-- 2120  
Db 2112 ---PQPPP-----PESCCALPPRSLPDPFSRPVPSQSSSQSPSLTPRPLSABAF 2161  
QY 2121 PENLVDSRGRSRGPKSPERSHVSSEPEPISP---POVPVVEKQDSILLILLS-QRGAEP 2176

```
Db 2162 PSPVTPRFQSPDYSRPPSRQSRDPFAPLHKPPRPQPPEVAFKAGSLAHTSLGAGGPPA 2221
QY 2177 -----BQRDARSQGISYLPSPFFTKLENTSPMVKSKQEIFRKLNS 2219
Db 2222 ALPAGPAGELHAKVPSGQPPNFVRSPGGAFVGT-----PSMRFTFPO----- 2265
QY 2220 GGGSDMAAAGPGEIFNLPAVTTSGSVSSRG-HSFADPASNLGLEDIIRKAL-MGSP-- 2275
Db 2266 -----AVGPELSK-----PPVQGLPPPHGINSHPGPTLGKPGQSTNVTATGNFHP 2314
QY 2276 -----DKVEDHGVVMQPGVVP-----GTANTSVVTSGETRREE-- 2311
Db 2315 SGSPGLPGSGTGESYGLSPLRPSPVLPAPDGLSPYLSHGASQSRGITSVPEKREDPG 2374
QY 2312 -----GDPSPHSGVCKPKL-----ISKNSRKS----- 2336
Db 2375 TGMGSSSLAELPQTQDPMGSLSQTELEKQORQRLRELIQRIQNTLRQEKETAAA 2434
QY 2337 --SPIPGGYLGTERRPSSSVHSEGDYHROTPCGMAWEDRPSSTGSTOFPYNPLTMRMLS 2394
Db 2435 AAGAVGPPGNGAPSSPAFQLSRG-----QTPAGTQDKSLVG---LPPSKLSGPILG 2487
QY 2395 -----STPPTPIACAPSAYN-----QAAPH-----QONRIWEREP 2424
Db 2488 PGSPFDDRLSRPPPPAPPSSMDVNSRLVGGSAFYORAPYPGSLPLOOQOOLWQOQ 2547
QY 2425 A 2425
Db 2548 A 2548

RESULT 45
JC5630
TCOF1 protein - mouse
C:Species: Mus musculus (house mouse)
C;Date: 14-Oct-1997 #sequence_revision 07-Nov-1997 #text_change 05-Nov-1999
C;Accession: JC5630
R:Paznekas, W.A.; Zhang, N.; Gridley, T.; Jabs, E.W.
Biochem. Biophys. Res. Commun. 238, 1-6, 1997
A:Title: Mouse TCOF1 is expressed widely, has motifs conserved in nucleolar phosphoprote
A:Reference number: JC5630; MUID:97445113
A:Accession: JC5630
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1320 <PAZ>
C:Cross-references: DDBJ:AF001794; NID:g2109458; PIDN:AAB71347.1; PID:g2109459
C;Comment: This protein is a nucleolar phosphoprotein with 82 potential phosphorylation
ing.

Query Match 2.6%; Score 334.5; DB 2; Length 1320;
Best Local Similarity 19.9%; Pred. No. 9.7e-06;
Matches 319; Conservative 197; Mismatches 610; Indels 481; Gaps 70;

QY 503 QQIARPSQEKVEKEEDKAEKTEKKEEKDEKEDKEDKENTKEKIDGTABETE 562
Db 54 QQTSELGQKQAEDETLQAKSRVSDPVSSESSDQKEE-----EAAT 98
QY 563 EREQATPR-----GR-KTANSQGRKKGRITRSMTEAAAAAASAAAAA 603
Db 99 ERAKATPRPTVNSATAALPSKVEKCKTKTAN-----KTVNSVSHPGSGKTVVHLLS 151
QY 604 TEPPPPPLPPPEPTSTPEVTSRWTEEME-----VAKKGLVEHGRNMAAI----- 650
Db 152 GKSP-----KKAEPPLANTVLAEE-----TEEGNAQALGPTAKSGTSAQSGSSSEDSISS 205
QY 651 -----AKMVGTKSEACKNFYNYK-----RRHNDLNLQHQKTKSRKPREERDVQ 698
Db 206 DETDVEVKSAPAKAQAASAPAKDPAPARTAPGPTKLGTV-----APTAPKAPAAAAA 260
QY 699 CESVASTYSAQDEEDIEASNEENPEDEVEAVKPSDENATSRGNTEPAVELEPTTE 758
```

```
Db 261 AAAVAAAAAAAE--SESEEDSDSEDEAPAGLPQ---VKASGKG---PHVRADSVSA 312
QY 759 TAPS-TSPSLAVP-STKPADES-VETQVNDISIAETAQMDVDQOEHSAEESGVCDDPPP 815
Db 313 KGISGKPIILATPKGTQPAATQAKAERPKDS---EFSSEDDSDSE----- 355
QY 816 ATKADSDVDEVVRVENHAS---KVEGDNTERD-LDRASEKVEPRDEDLVVAQIINAQR 870
Db 356 ---DEMPVIVNTQARTSGKSPRARGTSAPAKESSKGAPAVTPGKARPVAAQ---AGK 408
QY 871 PEPOSDNDSDATCSADEVDGEPERQRMFMDSKPSLLNPTGSLVSSPLKPNLDLPQL 930
Db 409 PEAKSSEES-----ESDSE-----TPAAATLTTSKPVKPLG--- 441
QY 931 QHRAAVTPMVVSCTP-----CNLPICTPVGYALYORHIKAMHESALLLEORQREQID 984
Db 442 --KSSQVRPVSTVTPGSGGKANLPCPGKVGSAALRVQMVK-----REDVSESSAE 491
QY 985 LECSRSTSPCGTSTKS---PNREWELQAPAPHLQITNLPPEGVRLPTTRTPPPPLIPSSK 1041
Db 492 LDSGPGSPAKAKASLALPKQ---VRPVATQVKTDRKGHSGSESSDSEEEAFAA- 546
QY 1042 TTVASERPSP-IMGGISISGTP---GTYLTSNQASYQTETPKPSVSGISLGLPQOE 1095
Db 547 -SAAQAKPALEKQMKASSRKGTPTASATCASTSSHCKAGAV--TSSASLSPALAKGTQ 603
QY 1096 SAKSATLPIYIKQEEFSPRSONSQPEGLLVRAQHEGVVGTAGAIQEGSITRGTPTSISV 1155
Db 604 DVDS-----SSESESE---AAPSTPRVQKSG----- 628
QY 1156 ESIPSLRGSITQGTTPALPOTGIPTALVKG-SISRMPIEDSSPEKGREEAASKGHVIEG 1214
Db 629 -----GKGLQKAAALQGQVAVVHTQKTGPSVKMAQEDS--ESLEEDSSSE- 672
QY 1215 KSHILSYDNIKNARECTRSPRTAHEISLARSYESVEGNKQGMs-----MRESVPSPAPLE 1270
Db 673 -----EEDETPAQATPL-----GRLPQAKANPPPTKTPPASA--S 705
QY 1271 GLICRALPRGSHDLKERTVLSGSIQOGTPRAITSEFEDGLKYPKQIKRESPPIRAFEG 1330
Db 706 GKAVAAATKGP-----PVPNSIV-SARGORSVPAAGKAGAPATQAKG 748
QY 1331 AITGKPYDGITTIKEMGRSIEHPIRQDILITQESRKTPVEVQSTRPIIE----- 1379
Db 749 PVA-GTGEDSESSKEESDEEETPAQ---IKPVGKTSQVRAASAPAKESPKKGAHPCTP 804
QY 1380 ---GSISQGTPIKFDNNSQSAIKHNKVSILITGPSKLSRGMPLIIVPENIKVVERGY 1435
Db 805 GKTSSATQAPQGTEDSDSSESDSDTEM---PSAQAIKSPVPVSNRNSSPVAPATPP 861
QY 1436 EDVKAGETVRSRHTSVVSSGSPSVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRG 1495
Db 862 EGQAVNTTKASGTTAQS-----SSEEDGEDLIPATQPSYALRTSVTTPAALSRA 916
QY 1496 SPMMNRSTDVITPPNKSTNIHERKSTLTPTQRESITPAKSPVGVDPVSHSPFDPHHRGST 1555
Db 917 ASOPSKSEQSRMPKGGKAKAASAQTSASAVETLPMMP-----QSAPIQ-----KA 964
QY 1556 AGEVYWHLP--TOLDPAMP-----FHRALDPAAAY 1595
Db 965 TNKLGKSKLPEKQOLAPGYPRAPRSSESDSDTSEDEDAKRPOMPKSAHR-LDPDPS-- 1021
QY 1586 LFORQ--LSPTPGPSQYQLYAMENTQTIINDYIT-----SQOMQVNLRPDVARGLS 1636
Db 1022 --OKETVVEETPTESSEDEWA---PSQSLLSGYMTPLGLIVANSQASKATPRD----- 1070
QY 1637 PREOPLGLPYATRGTIDLTNMPPTILVPHPGGTSTPPMDRITVYPTQTITPPRPVNSA 1696
Db 1071 --SNSLASSAPATKDNPDGKQSKS---QHAADATALPKTKRKEASSGSTPQPKKKKST 1125
QY 1697 SMSPGHPHTLAAAASAEER 1756
Db 1126 SSSPA-PTQILPNSITQRLLEQ-----AMPLSEAQVQA 1157
```

```

QY 1757 PSPSVRTOETMLQORPSVFOGTNGTSVITPLDPTAQLRIM--PLPAGGPSISQGLPASRY 1814
Db 1158 SVVKVLTE--LLEQ-----ERLKATEAKESGKKSQRKLSGDLGAP----- 1199
QY 1815 NTAADALAALYDAASAPQMDVSKTESKHEAARLEENLRSRRAVSEQ--QLEQKYLE 1872
Db 1200 -----KNKKKEQPV-----PRASAVSPEKAPMTSKAKSK 1229
QY 1873 VEKRSVQCLYTSSAPPSG---KPQPHSSSVYSEAGKDKGPPPKSRYEELRTRGKTTITA 1929
Db 1230 LDKGAGGKGGSPGQCAKEKPDGELLGKLESGEQSDPKSKKKKSLAKK----- 1282
QY 1930 ANFIDVIITROIADKDARE-----RGOSDSDSSSLSSHRHYETPSDA 1972
Db 1283 -----KDKKKKKKKKKSLAKDSASPIOKKKKKKKKSA 1316

```

Search completed: September 8, 2001, 14:44:04  
Job time: 426 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 14:34:18 ; Search time 34.32 Seconds  
(without alignments)  
4310.097 Million cell updates/sec

Title: US-09-522-753-11  
Perfect score: 12643  
Sequence: 1 MSSGGPPNQGAFSTEQSRY.....EREPAPLLSAQYETLSDSD 2440

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0601.\*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score   | Query Match % | Length | ID          | Description        |
|------------|---------|---------------|--------|-------------|--------------------|
| 1          | 11873.5 | 93.9          | 2343   | 21 AAB12453 | Human HNRCR protei |
| 2          | 11536.5 | 91.2          | 2453   | 21 AAB12454 | HNRCR protein sequ |
| 3          | 4168    | 33.0          | 2518   | 21 AAB40574 | Human ORFX ORF338  |
| 4          | 2885.5  | 22.8          | 619    | 17 AAR99738 | Retinoid X recepto |
| 5          | 1879    | 14.9          | 1495   | 18 AAW18226 | Transcriptional co |
| 6          | 520     | 4.1           | 3266   | 21 AAB42491 | Human ORFX ORF2255 |
| 7          | 507.5   | 4.0           | 2819   | 22 AAB35408 | Human 07CG27 gene  |
| 8          | 404     | 3.2           | 3256   | 21 AAV50976 | Human cell cycle p |
| 9          | 370     | 2.9           | 73     | 15 AAR53557 | Thyroid hormone re |
| 10         | 370     | 2.9           | 73     | 20 AAV40596 | Partial amino acid |
| 11         | 370     | 2.9           | 73     | 20 AAW92403 | Human TR-interacti |

|    |       |     |      |    |           |                     |
|----|-------|-----|------|----|-----------|---------------------|
| 12 | 370   | 2.9 | 73   | 20 | AAW85115  | Thyroid hormone re  |
| 13 | 368   | 2.9 | 2843 | 13 | AAR26052  | APC gene product 1  |
| 14 | 368   | 2.9 | 2843 | 18 | AAW35392  | Human adenomatous   |
| 15 | 368   | 2.9 | 2843 | 19 | AAW38370  | Human adenomatous   |
| 16 | 368   | 2.9 | 2860 | 15 | AAAR63507 | Adenomatous polyo   |
| 17 | 366   | 2.9 | 2843 | 15 | AAAR58634 | Adenomatous polyo   |
| 18 | 365   | 2.9 | 2742 | 21 | AAAR23012 | Human APC protein   |
| 19 | 365   | 2.9 | 2842 | 15 | AAAR63508 | Adenomatous polyo   |
| 20 | 365   | 2.9 | 2843 | 19 | AAW76140  | Human APC protein   |
| 21 | 365   | 2.9 | 2843 | 19 | AAW76144  | Human APC protein   |
| 22 | 365   | 2.9 | 2843 | 21 | AAAB23011 | Human APC protein   |
| 23 | 365   | 2.9 | 2973 | 19 | AAW76821  | Human APC protein   |
| 24 | 365   | 2.9 | 2973 | 21 | AAW70304  | Protein used in ca  |
| 25 | 365   | 2.9 | 2973 | 22 | AAW72782  | Transcriptional ac  |
| 26 | 364   | 2.9 | 2843 | 16 | AAW11922  | Adenomatous polyo   |
| 27 | 356.5 | 2.8 | 2781 | 21 | AAW57453  | Human transcriptio  |
| 28 | 349.5 | 2.8 | 3118 | 22 | AAW50362  | Human SRCAP. Homo   |
| 29 | 341.5 | 2.7 | 2907 | 21 | AAW57452  | Human transcriptio  |
| 30 | 329   | 2.6 | 2972 | 22 | AAW50363  | Human SRCAP. Homo   |
| 31 | 327   | 2.6 | 2783 | 13 | AAAR23963 | AFP-1 (Ala 2460 Va  |
| 32 | 327   | 2.6 | 2971 | 21 | AAAB1231  | Human ORFX ORF995   |
| 33 | 326   | 2.6 | 2783 | 13 | AAAR23962 | AFP-1. Homo sapie   |
| 34 | 325   | 2.6 | 1596 | 18 | AAW31347  | Rat tumour suppres  |
| 35 | 323.5 | 2.6 | 1780 | 19 | AAW53863  | Human gravin polyp  |
| 36 | 323.5 | 2.6 | 1780 | 21 | AAAB15380 | Human gravin prote  |
| 37 | 317.5 | 2.5 | 1341 | 21 | AAW85657  | Human Acinus L pro  |
| 38 | 316   | 2.5 | 3910 | 14 | AAAR38470 | ALL-1 protein. Ho   |
| 39 | 316   | 2.5 | 3910 | 16 | AAAR66462 | ALL-1 (acute lymph  |
| 40 | 313.5 | 2.5 | 1719 | 17 | AAAR92100 | Human RIZ allele D  |
| 41 | 311   | 2.5 | 1719 | 21 | AAAB12114 | Human mutant RIZ p  |
| 42 | 309.5 | 2.4 | 2441 | 16 | AAAR79054 | CREB binding prote  |
| 43 | 309.5 | 2.4 | 2441 | 21 | AAW94252  | Mouse nuclear CREB  |
| 44 | 309   | 2.4 | 1719 | 21 | AAAB12029 | Human RB-interacti  |
| 45 | 308.5 | 2.4 | 2441 | 19 | AAW40058  | Cellular transcript |

## ALIGNMENTS

|          |                                                                        |
|----------|------------------------------------------------------------------------|
| RESULT   | 1                                                                      |
| AAB12453 |                                                                        |
| ID       | AAB12453 standard; Protein; 2343 AA.                                   |
| XX       | AC AAB12453;                                                           |
| XX       | DT 24-OCT-2000 (first entry)                                           |
| XX       | DE Human HNRCR protein SEQ ID NO:20.                                   |
| XX       | Human; HNRCR; nuclear receptor coreceptor.                             |
| OS       | Homo sapiens.                                                          |
| XX       | PN CN1250094-A.                                                        |
| XX       | PD 12-APR-2000.                                                        |
| XX       | PF 06-OCT-1998; 98CN-0120919.                                          |
| XX       | PR 06-OCT-1998; 98CN-0120919.                                          |
| XX       | (XINH-) XINHUANGPU FUDAN GENE ENG CO LTD SHANGHA.                      |
| XX       | PI Yu L, Tu Q, Zhao Y;                                                 |
| XX       | DR WPI; 2000-400830/35.                                                |
| XX       | DR N-PSDB; AAA60629.                                                   |
| PT       | Preparation of new human keron acceptor co-repressor coding series and |
| XX       | the polypeptide -                                                      |
| PS       | Claim 4; Page 25-30; 58pp; Chinese.                                    |
| XX       |                                                                        |

CC The present sequence represents a human homologue of nuclear receptor  
CC coreceptor (HNRCR).

```

SQ Sequence 2343 AA;
Query Match 93.9%; Score 11873.5; DB 21; Length 2343;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 2316; Conservative 5; Mismatches 6; Indels 129; Gaps 5;

Qy 1 MSSSGYPNQAFSTEQSRYPHVSQVYTFPNTRHQOEFAVPDYSRSHLEVSAQSOLLQQQ 60
Dy 1 msssyppnqgafstegsrphvsqvytfpntnrhqoeafavpdysrshlevsaqsollqq 36
Qy 61 QQQQLRRRPSLLSEFPGSDRQPRERTSYEPFHPGCPSPVDHDSLESKRPRLEQVSDSHFQ 120
Dy 37 ----- 36

Qy 121 RVSAAVLPLVHPLPEGLRASADAKDPAGGKHEAPSSPSIGQPCGDDONASPSKLSKEE 180
Dy 37 -----dpafgkheapspsigqpcgddqnaspsklskee 71

Qy 181 LIQMDRVREIAKVEQILKLKKQQQLEBAAKPPPEKPPVSPVPVEQKHSIVQIYY 240
Dy 72 liqmdrvrelakveqilklkkqqgleeaaakppekppvsvppveqkhrslvqiyy 131

Qy 241 DENRKAEEAHKIPFGLGPKVELPLYNOPSDTKVYHENIKTNQVMRKKLILFFKRRNHAR 300
Dy 132 denrkaeeahkifeglpgkvelplynopsdtkvynenikcnqvmrkkllilffkrnhar 191

Qy 301 KOREQKICORYDQLMEAWKVKVDRIENPNRRKAKESKTREYYEQFPEIRKOREQOERFQ 360
Dy 192 kreqkicqrydqlmeawekvkvdrilennprnkakesktreyeqfpeirkreqqerfq 251

Qy 361 RVQBGAGLSATIARSEHEISFIDGLSEQENNEKQMRQLSVIPPMFMDAQRVKFINM 420
Dy 252 rvqrgaglsatiarseheisfidglsegesnekqmrqlsvippmmfdaeqrvkfnnm 311

Qy 421 NGLMEDPMKVYKDRFMVMTDHEKEIFKDFIQHPKNFGLIASYLEKSVPCDVLVYYL 480
Dy 312 nglmedpmkvdkdrfmvmtdhekeifkdfiqhpnkfgliasylerksvpdcvlyyl 371

Qy 481 TKKNENYKALVRNNGKRRGNQOIARPSQEEKVEEKDEKAETKEKEEKKEDEEKDE 540
Dy 372 tkknenykalvrnygkrrgnqoiarpsqeekveeekdekaetkekeekkeedeekde 431

Qy 541 KEDSKENTKERDKIDGTAETFEERQAPRGKRTANSOGRRKGRITRSMTNEFAAASAAA 600
Dy 432 kedskenkerdkidgtaeteerqacprgrktansgrrrkgritrsmtneaaaaaaa 491

Qy 601 AAATEEPPPPPLPPPEPTSTPEFVTSRWTEEMEYAKKGLVEHGRNMAAIAKMGVTKSEA 660
Dy 492 aaateepppplpppepistepvetsrwtееemevakkglvehgrnmaaiakmgvtkse 551

Qy 661 QCKNPFYKRRHNLNLLQHKOKTSKRPRERDVSCESVASTVSAQEDIEDIASNEE 720
Dy 552 qcknfykrrhnlldllqhkktskrpreardvscqsvastvsaqediedieasnee 611

Qy 721 ENPEDESE-----VEAVKPSDESPENATSRGNTPEPAVELEPTTETAPSTS 764
Dy 612 enpedesgaenssdtesapspsvavakpsdpsena csrgntpepavelepttetapsts 671

Qy 765 PSLAVPSTKPADESVETQVNDSTISAETAQMDVDDQOQSHSAEESVCDPPPPATKADSYDV 824
Dy 672 pslavpstkpaedesvetqvndstisaetaeqmdvdqqhseaeegsvcdpppatkadsydv 731

Qy 825 EVRVPENHASKVEGNTKTERDLDRASEKVEPRDEDLVVAQIINAQRPEPQSDNDSSATCS 884
Dy 732 evrvpenhaskevgn tkterdl drasekveprdedlvvaqinaqrpepqsndndssatcs 791

Qy 885 ADEDVDGEPERQMPMDSPKSLNPTGCSILVSSPLKPNPLDLQLOHRAAVIPPMVST 944
Dy 792 adedvdgeperqrmfmdskpslntptgcsilvssplkpnpldlqlqhraavippmvst 851
```

```

Qy 945 PCNIPITGTPVSGYALYQYRHKAMHESALLLEORORQEQIDLECRSSTSPCGTSKSPNREW 1004
Dy 852 pcnipitgtpvsgyalyrhikamhesalleeqrqeqidlecrsstspcgtskspnrew 911

Qy 1005 EVLQAPAPHLQILNLPEGVRLPTTRTPRPPPLIPSSRTTIVASEKPSFIMGSGISQGTPTG 1064
Dy 912 evlqapahqvltlnlpegvrlpttrtprrppplipsskttvasekpsfimsgsisqgtptg 971

Qy 1065 YLTSHNOASTYQETPKPSVSGISIGLPRQBSAKSAILPIYKQEEFSPRSONSOPEGLLV 1124
Dy 972 yltshnqasytqetpkpsvsgisiglprqbsaksailpiykgeefspresqnspegllv 1031

Qy 1125 RAOHEGVVGRGTAGAIQIAGSITRGTPTSKISVESITPSLRGSITQCTPALPQIGITEALVK 1184
Dy 1032 raqhegvvrgtagaigaiqegsitrgtptskisvesitpslrsgsitqctpalpqigitealvk 1091

Qy 1185 GSISRMPIEDSSPEKGREAAKSGHVIYEGKSGHILSYDNIKNAREGTRSPRTAHEISLK 1244
Dy 1092 gsisrmpiedsspekgreaaaskghviyegksgihlsydniknaregtrsrptaheislk 1151

Qy 1245 RSYESVGENIKQGMRESVPVAPLEGLICRALPRGSPHSDLKERTVLSGIMQGTPTAT 1304
Dy 1152 rsyesvegnikqgmresvpvsaapleglicralprgshsdlkertvlsqsimqgtprat 1211

Qy 1305 TESPEDGLKYPKQIKRESPPITRAFEAGATTGKPYDGTITTKEMGRSITHEIPRODILTQES 1364
Dy 1212 tesfedglkypkqikresppitrafegaitkgpydgtitkemgrsitheiprqdiltqes 1271

Qy 1365 RKTPEVVOSTRPIIEGSIISOGTPIKFDNNSGSAIKUNVKSILITGPKSLSGMPLEIVP 1424
Dy 1272 rktpevvqstrpiiegsiisogtpikfdnnsqsaikhnvksiligtgpkslsgmpleivp 1331

Qy 1425 ENIKVVERGKYEDVKAGETVRSRHTSVVSGPSVLRSTLHEAPKAQLSPGIYDDTSARRT 1484
Dy 1332 enikvvergkyedvkagetrsvrhtsvvsgpsvlrstlheapkaqlspgiyddsarrt 1391

Qy 1485 PVSYQNTMSRGSPPMNRSTSDVTIPPNKSTNHERKSTLTPTQRESIPAKSPVPGVDPVVSH 1544
Dy 1392 pvsyqntmsrgsppmnrtsdvs--snkstnherkstitptqresipakspvpgvdpvsh 1449

Qy 1545 SPDPHHRGSTAGVYVWSHLPTQLDPAMPFHRALDPAALAAAYLFORQLSPPTGYFSQQLY 1604
Dy 1450 spdphhrgstagvyrshlptqldpampfhraldpaalaaaylforqlspptgyfsqqlly 1508

Qy 1605 AMENTROTILNDYITTSQOMOVNLRPDVARGLSPREQPLGLPYPATRGIIDLNTNPPPTLV 1664
Dy 1509 amentrotilndyitsqomovnlrpdvarglspreqplglpypatrgiidlntnppptlv 1568

Qy 1665 PHPGGTSTPPMDRITYIPGTQITFPPRPNYSASMSGHPTHAAAAAERERERERERER 1724
Dy 1569 phpggstppmdrityipgtqitfpprpnysasmsghpthlaaaaaerererereker 1628

Qy 1725 EREERIAAASDLYLRPGEQGRPGSHGYVRSPSVYRTQETMLQORPSVPOGTNGTSVI 1784
Dy 1629 ereriaaasdlylrpgeqgrpgshgyvrspsvyrtqetmlqorpsvpoqngtsvi 1688

Qy 1785 TPLDPTAQLRIMPLPAGGPSISQGLPASRYNTAADAALAAALVDAASAAPQMDVSKTKESKH 1844
Dy 1689 tpldptaqlrimplpaggpsisqglpasryntaadaalaaalvdaasaapqmdvsktkesh 1748

Qy 1845 EAARLEENLRSRSAVSEQOOLEKTLVEKRVSQCLYTSSAFPSGKPPQPHSSVYSEAG 1904
Dy 1749 eaarleenlrsvsaavseqqlektlvekrvsqclytssafpsgkppqphssvvyseag 1808

Qy 1905 KDKGPPKPSRYEELRTRGKTTITAAANFIDVITITROJASDKDARERGSSDSSSSLSH 1964
Dy 1809 kdkgppkpsryeelrtrgkttitaanfdivititrojasdkdarergssdssssls 1868

Qy 1965 RYETPSDAIEVISPASSAPPQEKLOTQYQPEVVKANAENDPTRQYEGPLHHYRPPQESP 2024
Dy 1869 ryetpsdaievispassappqeklotryqpevvkanaendptrqyegplhhyrppqesp 1928
```

QY 2025 SPQOOLPSSQAEGQGVPRTHRLITLADHICQIITODFARNQVSSOTPOQPTSTFQNS 2084  
 Db 1929 spqgqlpssqaegmgvprthrilitladhicqitqdfarnqvssqtpqgptstfqn 1988  
 QY 2085 PSALVSTPVRKTSNRYSPEQAQSVHHQPGSRVSPENLVYDKSRGSPGKSPERSHVSS 2144  
 Db 1989 psalvstpvrtktsnryspesqaqsvhhqprgsrvspenlvdkrgsrpgkspershvss 2048  
 QY 2145 EPEYEPISPPQVPVYVHEKQDSLLLSQSGAEPAPQORNDARSPGISYILPSFPTKLENTSPM 2204  
 Db 2049 epeyepisppqvpyvvhkqdslllsqsgaepaqrndarspgsisylpsfftklentspm 2108  
 QY 2205 VKSKQEIFRKLNSGGGSDMAAAQCGTEIFNLPVAVTSGSVSSRGHSFADPASNIGLE 2264  
 Db 2109 vkskqelfrklnsgggdsmaaaqcgteifnlpavtsgsvssrgshsfadpsasnigle 2168  
 QY 2265 DIIRKALMGSDDDKVEDHGVVMSOPMGVPGTANTSVVTSGETRREGDPSPHSGGCKP 2324  
 Db 2169 diirkalmgstddkvedhgvmsopmgvpgtantsvvtsetreegdpshs-gvckp 2227  
 QY 2325 KLIKSNSRKSPIPGQYLGTERPSVSVHSEGYHROTQGWAWEDRPSSTGSTQFP 2384  
 Db 2228 kliksnsrkskipggyylgterpsvsvhsegyhrotqgwawedrpsstgstqfp 2287  
 QY 2385 YNPITMRMLSTPTPTIACAPSAYNAAPHOQNRWIEREPAPLLSAQYETLSDD 2440  
 Db 2288 ynpitmrmlstptptiacapsaynaaphoqnriwerepapllsaqyetlsd 2343

## RESULT 2

AAB12454  
 ID AAB12454 standard; Protein; 2453 AA.

AC AAB12454;

DT 24-OCT-2000 (first entry)

XX HNRCR protein sequence.

XX Human; HNRCR; nuclear receptor coreceptor.

XX Unidentified.

XX CN1250094-A.

XX 12-APR-2000.

XX 06-OCT-1998; 98CN-0120919.

XX 06-OCT-1998; 98CN-0120919.

XX (XINH-) XINHUANGPU FUDAN GENE ENG CO LTD SHANGHA.

XX Yu L, Tu Q, Zhao Y;

XX WPI; 2000-400830/35.

XX N-PSDB; AAA60630.

XX Preparation of new human kernon acceptor co-repressor coding series and

XX the polypeptide -

XX Example 2; Fig 2; 58pp; Chinese.

XX The present invention describes a human homologue of nuclear receptor

XX coreceptor (HNRCR). The present sequence represents an HNRCR protein

XX sequence used in comparison with the human HNRCR.

XX SQ Sequence 2453 AA;

Query Match 91.2%; Score 11536.5; DB 21; Length 2453;

Best Local Similarity 91.0%; Pred. No. 0;

Matches 2245; Conservative 79; Mismatches 101; Indels 43; Gaps 12;

QY 1 MSSGYPPNOCAGFSTEQSRYPHPSVQYTFPNTRHQEQFAVDPYRSSHLEVSQAQLQOQ 60  
 Db 1 mssgyppnccgafstegrypsvshvqytfpnsrhaqefavpdyrsshlevsaqslqqq 60  
 QY 61 QQQQLRRRPSLLSEFFHFGSDRPERRTSYEPFHPGSPVDHDSLESKRPRLEQVSDSHFQ 120  
 Db 61 qqqlrrrpsllsefhpgrsdprerrsgyefhpgspvdhdsleskrprleqvdsdshf 120  
 QY 121 RVSAVLPLVHPLPEGLIRASADAKDPAGCKHEAPSSPISGQPCGDDONASPKLSKEE 180  
 Db 121 risaavlpvlhplpeglirsaanakkdpagvckheapspslgpcgddqnaspsklskee 180  
 QY 181 LIQMDRVDRREIAKVEQOILKLLKKQOOLEEAAKPEPEKPPVPPVPEQHRSTIVQIY 240  
 Db 181 liqmdrvdrreiakevqilklkkqqleeeaaakppekpvspvppeqhrstivqiy 240  
 QY 241 DENKKAEEAHKIPGEGIPKVELPLYNQSDTKVYHNINIKTNQVMRKLILIFFRRNHAR 300  
 Db 241 denrkkaeeahkifeglpkvelplynqsdtkvyniktnqvmrkliliffrrnhar 300  
 QY 301 KOREQKICORYDOLMEAEWEKKVDRIENNPRAKESKTREYVEKQFPEIRKQEOERFQ 360  
 Db 301 kreqkicorydqlmeawekkvdrriennprrkakesktreyyekqfpeirkreqqerfq 360  
 QY 361 RVGQAGLGSATIAARSEHEISEIIDGLSEONNEKQMRQLSVIPPMFADQRVVKFINM 420  
 Db 361 rvqrgaglsatiarseheiseiiddglseqennekqmrqlsvippmfdaeqrrvkvfinm 420  
 QY 421 NGLMEDPMKVKYKQFMVNTDHEKEIFKDFIQHPKNFGLIASYLERKSPVDCVLYYL 480  
 Db 421 nglmedpmkvykdrqfmnvntdhekeifkdfiqhpknfgliasylerksvdcvlyyl 480  
 QY 481 TKNKNYKALVRNRYGKRGNNQIARPSQEEKVEEKEEDKAEKTEKKEEKDEEKDE 540  
 Db 481 tknknycalvrnrygkrngnqiarpseqeekeekdaektekeekdkdeekdd 540  
 QY 541 KEDSKENTKEKDKIDGTAETEEREQATPRGKTANSQGRKRGRITRSMTNEAAAAA 600  
 Db 541 kedskentkekdkidgtaeteereqatprgkktansqgrkgvtrtsmtneaaaaa 600  
 QY 601 AAATEEPPLPPPPPEPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTP 660  
 Db 600 aaateepplpppppepistpistpistpistpistpistpistpistpistpistpistpist 660  
 QY 661 QCKNFYNYKRRHNLNLQOHHKQKTSRKPREERDVQCESVASTVSAQEDDIEASNEE 720  
 Db 660 qcknfyynykrrhnlldlqhhkqksrkpreerdvqcesvastvsaqededieasnee 720  
 QY 721 ENPEDSE-----VEAVKPSDESDPENATSRGNTPEPAVELEPTTAPSTS 764  
 Db 720 enpedsegaenssdtesapspsveaakssedssenaasrgntpepaeleatdpapcas 779  
 QY 765 PSIAVSTKPAEDSVETQVNDISIAETAEQMDVQOQDQHEAEGSVCDPPPATKADSVY 824  
 Db 760 pssavptkpaeresveaqvtdsasaetaepmdvdeecgaegsvldppaptkadvdp 839  
 QY 825 EVRYPENHASKVECDNTKERDLDRASEKVEPRDEDLVVAQOINAQRPQSDNDSSATCS 884  
 Db 840 emqvptentaskgegd-akerdlestsekteardedvvvaeqi--erpepqsdssatcs 896  
 QY 885 ADEDVDCGEPRQRPFMDSKPSLNPFGSLVSSPLKPNPLDLPLQHRAAVIPPVYVST 944  
 Db 897 adegdvgeprqrvfmdakpslltpgssllissipknlldlpqlhraavipmvst 956  
 QY 945 PCNIPITGVSGYALYQRHIKAMHESALLEQRQORQEQIDLECRSSTSPGCTSKSPNREW 1004  
 Db 957 pcnipitgvsgyalqyrhikamhesalleqrqrqeqvdlccrscstspcstskspnrew 1016  
 QY 1005 EVLQAPAPHQLITNLPEGVRLPTTRPPPLIPSSKTTTVASEKPSFIMGSIQGTPT 1064  
 Db 1017 evlqapahqvitnlpegvrlpttrtrppplipsskttvasekpsfimgssiqgtpt 1076

QY 1065 YLTSHNQASYTQETPKPSVSGSISLGLPQROBSAKSATLPIYIKOEFSPRSQNSQEGILV 1124  
Db 1077 yllshnda-yqeapkpqsvgsisglprqgestkaaplyikqeeefprsqnsqegilv 1135  
QY 1125 RAQHEGVVGRGAGAIQBSITRGPNTSKISVESIPSRGSIQTQTPALPQGTGIPTEALVK 1184  
Db 1136 raqhegvvrgtagavqegsitrgtpaaskisvetissirgsitqgtpalpqaqiptealvk 1195  
QY 1185 GSISRMPTEDSSPKGREAAKGHVLYEGSKGSHLSVDNIKNAREGTRSPRTAHEISLK 1244  
Db 1196 gvsirmlleespekvreeaasknlyegskgnilyndniknaregtrsrprtanemsik 1255  
QY 1245 RSYESVEGNIKQGMRESPPVAPLEGLICRALPRGSPHSDLKERTVLGSGIMQTPRAT 1304  
Db 1256 rsyaeavgsikqgmsmresppvapeleglicralprgspghsdlkertvlgsgimgtprat 1315  
QY 1305 TESFEDGLKYPKQIKRSPPTIRAFEGAITKGPYDGTITTKEMGRSHEIPRQDILTOES 1364  
Db 1316 aesfedgklypkqikresppirafegaikgkpydgittikemgrsheiprqdiltqes 1375  
QY 1365 RKTPEVQOSTRIIEGTSIQGTPIKFDNNSGOSAIKHNKSLITGPSKLSRGMPPLEIVP 1424  
Db 1376 rktpevqgstriieegisqgtpikfdnmsqsaikhnvkslietgyskplprgm--leivp 1433  
QY 1425 ENIKVVERGKYEDYKAGETVRSRHTSVVSGPSVLSRTLHEAPKAQLSPGYIDTTSARRT 1484  
Db 1434 enikvvergkyedvkagevtrvrshtsvvsgpsvlsrtlheapkaqlspglyiddssarrt 1493  
QY 1485 PVSQNTMSRSGPMNRTSDVTIPNKSHTNHERKSTLTPTQRESIPAKSPVPGVDPVVSH 1544  
Db 1494 pvsqntisrsgpmnmrtsdvs--ssksashekrstltptqresipakspvpgvdpvlsh 1551  
QY 1545 SPFPDHRHGSPAGEVYVSHLPTQLDPAMPFHRALDPAAAYLFQRLSPPTGYPQSQQLY 1604  
Db 1552 spfdphrssaagevyrshlptldpampfhraldp-aaayllqrqlspptgypsqqly 1610  
QY 1605 AMENTRQILNDYITSQOMQVNLRPDVARGLSPREQLGLGYPATRGLIDLTNNPPTILV 1664  
Db 1611 amentrqilndyitsqomqvnlpdpvtrglspreqplglpypatrglidltnnpptilv 1670  
QY 1665 PHPGGTSTPPMDRTYITPGTOITPPRPYNSASMSGPHPTHLAAAASAEEREREREK-- 1722  
Db 1671 phpggtstppmdrtityipgtqvtlpprpynaaslsphnpthlaaaasererereker 1730  
QY 1723 -----ERERERIAAASSDLRPSGEQGRPGSGHYVRSPSPSVRTQBTMLQQRPSV 1774  
Db 1731 ererererereraiaapadlylrpsgeqgrpgsghyvrspsvrtqbtliqqrpsv 1790  
QY 1775 FQGTNGTSVITPLDPTQALRIMPLPAGGPTISQGLPASRYNTAADALAAALVDAASAPQM 1834  
Db 1791 fqgtngtsvitpldptqalrimplpsgppsissqglpasryntaadalaalvdaasapqm 1850  
QY 1835 DVSKTKESKHEAARLEENLRSRAVSEQOOLEKTEVEKRSVQCLYTSSAFPSGKQPQ 1894  
Db 1851 dvsktkeskheaarleennlrsraavseqqglekntevekrsvqcltssafpsgkqap 1910  
QY 1895 HSSVYVSAGDKGPPPKRSYEELRGTGTTITAAFNIDVIIITRQIASOKDAREGSSQS 1954  
Db 1911 hasvvyseagdkpppkrsyeelrtrgtttitaanfivdiitrqiasdkdaregssqs 1970  
QY 1955 SDSSSSLSHRYETPSDAIEVISPASSAPPQEKLOQYQPEVVKANQAEENDPTROYEGPL 2014  
Db 1971 sdsssslsshrjetasdaievispassappqekpqqayqpdmvkandqaenestroyegpl 2030  
QY 2015 HHYRPOQESPSPOQ--LPPSSQAEQMGQVPRTHRLTLADHICQIITQDFARNQVSSQT 2072  
Db 2031 hhyrqqespsppqqplppssqegmgqvpvrthrltladhicqilqtqdfarnqv---- 2086  
QY 2073 PQQPPTFTQNSALVSTPVRTKTSNRYSPESQASVHHQRPGRSPVSPENLVKSRGRS 2132  
Db 2087 psqaststfqtspasaltstprtktsrnysspesqsvlhpvrprvspenlvksrgrs 2146  
QY 2133 PGKSPERSHVSSEPEYIPSPQVPVHVKQDSLILLSLQSGAEPAEQNRDARKSPGISYLP 2192

Db 2147 pgkspershipsepyepispqpqpavhekqdsmlllsqrgvdpaeqrsdsrpsgsisylp 2206  
QY 2193 SFFTKLENTSPWVKSQKQEIFRKLNSGGGSDMAAAQPGTEIFNLPAVTTSGSVSRGH 2252  
Db 2207 sfttklestspmvkskqgeifrlnssgggdsmaaaqpgteifnlpavttsgavsrsh 2266  
QY 2253 SFADPASNLGLEDIIRKALMGSDKVEDHGVMQPMGVVPGTANTSVVTSQGTREEG 2312  
Db 2267 sfadpasnlglediirkalmgssfdkvedhgvmshpvgimpssastsvtssearrdeg 2326  
QY 2313 DPSPHSGGVCKPKLISKSNSRKSPIPGQGYLGTTERPSSVSVHSEGDYHROTPGWAVE 2372  
Db 2327 epspha-gvckpklinsksrkskspipggsyigterpsvsvhsegdyhrqtpgwawe 2385  
QY 2373 DRSSSTGSTOFFPNPLTMRMLSSSTPTPIACAPSAVNAQAAPHQONRIWEREPAPLLSAQY 2432  
Db 2386 drpsstgstoffpnpltrmlsstptpiacapsavnaqaaphqonriwerrepapllsaqy 2445  
QY 2433 ETLSDDSD 2440  
Db 2446 etlsdsdd 2453  
RESULT 3  
AAB40574  
ID AAB40574 standard; Protein; 2518 AA.  
XX AAB40574;  
AC AAB40574;  
XX 08-FEB-2001 (first entry)  
XX Human ORFX ORF338 polypeptide sequence SEQ ID NO:676.  
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
XX vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;  
XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;  
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
XX cholesterol ester storage; systemic lupus erythematosus; infection;  
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;  
XX thrombosis; contraceptive.  
XX Homo sapiens.  
XX WO200058473-A2.  
XX 05-OCT-2000.  
XX 31-MAR-2000; 2000WO-US08621.  
XX 31-MAR-1999; 99US-0127607.  
XX 02-APR-1999; 99US-0127636.  
XX 05-APR-1999; 99US-0127728.  
XX 30-MAR-2000; 2000US-0340763.  
XX (CURA-) CURAGEN CORP.  
XX Shimkets RA, Leach M;  
XX WPI; 2000-602362/57.  
XX N-PSDB; AAC74783.  
XX Novel nucleic acids and peptides derived from open reading frame X,  
XX useful for treating e.g. cancers, proliferative disorders,  
XX neurodegenerative disorders and cardiovascular disease -

Claim 11; Page 784-790; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnerary; antiproliferative; antiparkinsonian; nootropic; neuroprotective; osteoprotective; anticonvulsant; antithyroidic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, asthma, bacterial or fungal infection, malaria, autoimmune disorders, viral, allergic, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 2518 AA;

Query Match 33.08; Score 4168; DB 21; Length 2518;

Best Local Similarity 40.88; Pred. No. 1.4e-219;

Matches 1101; Conservative 340; Mismatches 784; Indels 476; Gaps 108;

17 QSRYPHSHVQVTFPTNRHQEFAPVDYRSSHLEVSQASLQOQQOQLRRRPSILSEFH 76  
 16 eprypshslsvpviarhtdvgileq- hhsrdyashlspgslqprrlpallef 73  
 77 PGSDRQD- -RTSYEPHFGPSVPVHDSDLESKRPRLEQVSDSHFQRVAALVPLHPL 134  
 74 pgnersqelhlrpeshsylpelgkxsefieskrprlelpld- - - - -pllrpsp 122  
 135 ---EGLRA-SADAKDPAFGKGKHEAPSSPTSGQPCGDONASPSKLSKEELIQSDMDVRD 190  
 123 llatqpagseelldkdrsltgkle-pvsppsphtdpelelvprlskeelilqmdrvdr 181  
 191 EIAKVEQOILKKKKQOOLEEAAKPEPEKPVSPVPEQKHSRIVQIYDENRKAEEA 250  
 182 eitmvqgisklkkqgleeaaakpepekpvsppleshrslvqilydenrkkaeaa 241  
 251 HKIFEGLPKVELPLYNQPSDTKYHENIKTNQVWRKLLIFFRRNRHARKQROKTCQR 310  
 242 hrilegipqvelplynqpsdtrqyhenikinqamrkklilyfrrnrharkqwekfqr 301  
 311 YDOLMEAWKVVDRINNPRKAKESKTRREYKQFPEIRKQEQOERFQ-RVGORGAGL 369  
 302 ydqlmeawekvriennprrrakesvreyekqfpeirkqelqermgvrvgqrgsl 361  
 370 SATIARSEHETSEIDGLSEQENNEKQRLSVIPPMFADAEQRRVFINNGLMEPMK 429  
 362 smsaarsehsevlidglseqenlekqmrqlavipmlydaddqgrikfinmnglmadpmk 421  
 430 VYKQRPQWVWTDREKEIFDKFTQHPKNFGLIASYLERKSVPCVLYYYITKKNENYKA 489  
 422 vykdrqvmnmwsekeketfekfmqbpknfgliasflerktvaecvlyyyitkknnyks 481  
 490 LVRRNYKRRGRNOQIARPSQEEKVEEKEED- - -KAEKTEKKEEKDEEKEDEKESKE 546  
 482 lvrrys-rrrgsqqqqqqqqqqqqqqqqqmprrsqeekekekeakeeekpeve 540  
 547 NTKE- - -KDKIDGTA-BETEEREQATPRGRKTANSQGRKGRITRSMTNEAAAAASAAAA 602  
 541 ndkedlllkektddsgndekaeavaagkrktansqgrkgritrsmaneanseaatpq 600  
 603 ATEEPPLPPPPPPPISTEPVTSRTEEMEVAAGKGLVEHGRNWAIAIAKVMGKSAQC 662

601 qs-----aelasmelnessrteeemetakkgllleghrnwsaiarnvsgktvsqc 650  
 663 KNFYFNKRRRNHLNLLQOHKQKTSRKPRERDVSQCESVASTVSA- - -OEDEDEIAS- 717  
 651 knfyfnykkrqnldeilqghklmekernarrkkkapaaseeaaafppvvdeemeasg 710  
 718 ---NEEENPEDE- - - - -VEAVK- - - - - 732  
 711 vsgneemveeaealhasgnevprgecspatvnnssdtesipshteaaakdtgngkpk 770  
 733 -----PSEDSPEPATSGNTEPAVELE- - -PTTETAPSTSPSLAVPSTKPADE 778  
 771 patlgadgppprrprtrrptaseatgaptppap-psps-appvvvypvke 828  
 779 SVETQVNDISIAETAEQMDVQDQSHSAEESVCDPPPPATK- - -ADSDVDEVRVPENHASKV 836  
 829 eeetaaappv- - - - -eegeekppaaeelaavdtgkaeepvksecteaa 871  
 837 EGDNTKERDLDR- - - - -SEKVEPRDELVVAQOINQAQRPPEPQSDNDNSATCSADE 887  
 872 eegpakgkdaaeaaeagalkaekkggrattaksga- - - - -pg-dsdsatcsade 926  
 888 --DVDGEPERMPMDSKPSLLNPTGSLV-SSPLKPNPLDLPLQHQHRAAVIPPMWVSC 944  
 927 vdeaeeggdknrlsp- - -rpsalltptgdpranaqpk- - -pidlklqkqraaaippi- - - 977  
 945 PCNIPIGTPVSGYALYQRIKAMHESALLE- - - - -EORQROEQIDLECRSSTS 992  
 978 -----qvtkvheppredaaptkppappppppqnlqpesdapqpgss 1019  
 993 PCGTSKSP- - -NRE- - - - -W- - - - -EVLQAPAP- - - - -QLINFL 1018  
 1020 prgksrppadkaeataeaaqklpgdpcwtsglfpvpprevikaspdpasafaya 1079  
 1019 PEGVRLP- - - - -TRPRRPPPLPSSKTTVASEK-PSFI- - -MGSISQGTGTYLTSHN 1070  
 1080 ppgphlplghltdarvlprrpdtisnppplissakhpavlerqigalsqg- - - - -msvql 1134  
 1071 QASVTQETPKPSVSGISILGLPRQESAKSATLPYIKOEFPSPSONSQPEGLLV-RAQHE 1129  
 1135 hvpysehakap-vpvtmgllpmdpkklapfsgvkeqlsprgagppeslgvptaea 1193  
 1130 GVVRGTA-GAIOEGSITRGTTFTSKISVESIPSLRSGITQGTTPALPQTGIPTEALVKGSGIS 1188  
 1194 svlrgtalgsyvggsitkgipstrvpsdsaityrgsichgtpa- - - - -dvlykgkit 1245  
 1189 RMPEDSSP- - -EKGREAAAKGHVIECKSCHILSYDNKNA- - - - -REGTRSPRTAHEI 1241  
 1246 rilgedspsrldrgedsipkghviyegkghvisyeggmstvtqcskedgrssagpphet 1305  
 1242 SL-KRSVESVEGNIKQGMRESVPVAPLEGLICRALP- - -RGSPHSDLKERTVLSGSTMQ 1298  
 1306 aaprttydmmeqrvgrais- - - - -sasieglmraippperhshp-hlkeqghirgsitq 1358  
 1299 GTPRATSEFEDGL-KYPKQIKRES- - - - -PPIRAFEGAI- - - - -TKGKP-YDG-ITT 1343  
 1359 giprsyveaqedyirreakllkregtppppps-dltaeyktqalplklkpaheglvat 1418  
 1344 IKEMGRSHEIHPRODILTOESRKTPVVOSTRPIESISOGTPIKFDNN-SGQSAIKHN 1402  
 1419 vkeagrsiheipreel- - - - -rhtplaplprlkgsitgdtgtpikdytgastgskhd 1473  
 1403 VKSLITGPSKLSRCMPPLEIIVPENIKVVVERKGYEDVAGETVRSRHTSVVSGSGSVLRST 1462  
 1474 vrsliagspgrtfppvhpldvmad-aralaracye- - - - -eslkarpgtasssggsiarga 1527  
 1463 ---LHEAPKAQLSPGIYDDTSARRTPVSQNTMGRSGSPMMNRTSDVTIP- - - - -PNKSTNH 1515  
 1528 pviyelpgkpkqspltyedhga- - - - -pfaghlprgsptvtreptprlqegslssksaq 1582  
 1516 ERKSTLPTQRESIPAKSPVGPVDPVSH- - -SPDPHHRGSTAGEVYVSHLPTQLDP-AM 1572  
 1583 drkitstpre- - - - -iaksphstvpbehphpsihllrgvsgvdlyrshiplafadptsi 1638

```
Qy 1573 PFHRALDPAAYLFRQLSPTPGYQYLY-----AMENTROIINDYITSOQM 1623
Dy 1639 prgpld-aaayylprhlaptyphlypylrygypdtaalen-rqtiindyitsqgm 1696
Qy 1624 QVNL-----RPDVARGLSPREQPLGLDYPYA-TRGIIDLTMWPP-TILVPHPGTSTPPM 1675
Dy 1697 hntatamaqradmlrglspressialnyaagprgildsqvhlplvlpptpgtptam 1756
Qy 1676 DRIVYPTQTTFPRPNYSMSPGHPTHL---AAAASAEERERERERER---I 1729
Dy 1757 drlaylptcapfser-hssspispggthtkpttssssrerdrdrdrereksi 1815
Qy 1730 AAASDL-----YLRPGSBQ-----PCPGSHGYVRSPSP-SYRTOETMLQQ 1770
Dy 1816 lttsttvehaplrpgteqssgssgssgssarpashahqhsprtda-lq 1874
Qy 1771 RPSVFOGTNGTSVITPLDPTAQLRIMPLPAGGPSISQGLPASRYNTAADA-LAALVDAAA 1829
Dy 1875 rpsvlnhtgmkgiitavepstptvtrststspv-----rpaatfpapthcplgltldgy 1930
Qy 1830 SAPQMDVSKTESKHEAARLEENLRSAAVSEQOOLEQKLEVEKRSVOCILYTSAPPS 1889
Dy 1931 ptlmepvllpk-----eaprvaperprad-----tghafila 1962
Qy 1890 GKP-----QPHSSVYVYAGDKGP--PPKSYREELTRGKTITTAANFIDVITRQIA 1942
Dy 1963 kparsgilepass---psksepplvpvshgatiartpakn--laphhaspdpappa 2017
Qy 1943 SDKDAREGSSOSSDSS-----SSLSHRYETPSDAIEVISPASSPAPQKLOTYQDEV 1996
Dy 2018 sasphrektqsfisqlelrlgyhgsyspgevpvpspslthdkgipkhllee 2077
Qy 1997 VKANQAEENDPTROVEGP-----LHRYRP---QOESPSPQOOLPPSSAEGMGVPR 2044
Dy 2078 ldkshlegelrkpgqpvkvlgeaahphlrlpesqssplltqapg-----vkg 2129
Qy 2045 THRLLTADHCIIQITQDFARNOVSSOTPOQ---PPTSTFONSPSALVSTVP---RTKTS 2098
Dy 2130 hqvvtlaqhisevitqytrhh-----pqqlsaplpaplysfga---scpvldlrpps 2182
Qy 2099 NRYSPESQAQSVHQRGSRVSPENLVDKSRGSRGKSPERSHV---SSEPYEPISPPO- 2154
Dy 2183 dlylpppd-----hgap-argshp-----seggrkrspepnktsvlggedgievpspeg 2231
Qy 2155 -VPVVHEKQDILLLSQGAEPAPAEORNDARGPGSTSYLPSFETKL-ENTSPMWKSKQEI 2212
Dy 2232 mtephrsavypillyrdgedteprmgskspgntsqppaifskltsensamvskkqei 2291
Qy 2213 FRKLNSGGGSDMAAAQPGTEIFNLPAVTTSGVSSRGHSFADPAS-NLGLIEDIIRKAL 2271
Dy 2292 nklthnrnepeynisqpgteifnmpaitgtlmyrsqavqehastnmgleairkal 2351
Qy 2272 MGSFDDKVEDHGVVMSQPMGV---VPGTANTSV-----VTSGETRRREGDPSHSGGVCK 2323
Dy 2352 mgykdwqee-----spplsanafnlnasaslpaaampaadgrsdhltspgggg--k 2403
Qy 2324 PKLISKNSRKSQKPIPGQYGLGTERPSSVSSVHSEGDYHROTTP--GWAWEDRPSSTGST 2381
Dy 2404 akvsgprsrkaksap--glasdrpssvssvhssegdcnrrtpltvnrwvdrpsagst 2461
Qy 2382 QFPYNPLTMRM---LSSTPTPTIACAPSAVNQAAPHQONRIWEREPAPLLSAQYETLSDS 2438
Dy 2462 pfypnplmrlqagvmasppppgplpagsg-plagph---hawdeepklpscyetlids 2517
Qy 2439 D 2439
Dy 2518 e 2518
```

RESULT 4

AAR99738

ID AAR99738 standard; Protein; 619 AA.

```
XX AAR99738;
AC 27-SEP-1996 (first entry)
DT Retinoid X receptor interacting protein R1P13.
DE Retinoid X receptor interacting protein; RXR; RIP; R1P13.
KW Mus sp.
OS W09621677-A1.
PN 18-JUL-1996.
PD 08-DEC-1995; 95WO-US16311.
PF 13-JAN-1995; 95US-0372652.
PR (GEHO) GEN HOSPITAL CORP.
XX Choi H, Moore D, Seol W;
XX WPI; 1996-342241/34.
XX N-PSDB; AAT31931.
XX Retinoid X receptor (RXR) interacting protein (RIP) - useful to
XX modulate or mediate RXR function, anti-RIP antibodies can be used to
XX determine RIP subcellular distribution patterns
XX Claim 2; Page 51-52; 90pp; English.
XX Mouse retinoid X receptor (RXR) interacting protein R1P13 (AAR99738)
XX is a candidate transcriptional co-activator. It was identified using
XX an in vivo interaction trap system for the isolation of proteins that
XX physically interact with RXRs, esp. with the ligand binding domain of
XX human RXR alpha. Recombinant R1P13 can be obt'd. using a cDNA clone
XX (AAT31931) obt'd. from a mouse liver library. R1P3 (see also AAR99735-37
XX and AAR99739) can be used to modulate or mediate RXR function, and
XX may be used therapeutically or to raise antibodies.
XX Sequence 619 AA;
Query Match 22.8%; Score 2885.5; DB 17; Length 619;
Best Local Similarity 90.1%; Pred. No. 3.8e-150;
Matches 562; Conservative 27; Mismatches 28; Indels 7; Gaps 3;
Qy 1819 DALAALVDAASAFQMDVSKTESKHEAARLEENLRSAAVSEQOOLEQKLEVEKRSV 1878
Dy 1 dalaalvdaasapqmdvsktkskheaarlelenlrssaavseqqlqknlekrsv 60
Qy 1879 QCLYTSAFSGKPPQHSSVYVYSEAGDKGPPKSRVEEELTRGKTITTAANFIDVIT 1938
Dy 61 qcvctssalpsgkagphasvvyvysagkdkgppksryeeeltrgkttitaanfdivcit 120
Qy 1939 ROIASDKDAREGSSQSDSSSLSHRYETPSDAIEVISPASSPAPQKLOTYQPEVVK 1998
Dy 121 rqlasdkdaregssqsdssslshryetasdaievispasspappqkqayqpdnvk 180
Qy 1999 ANOENDPTROYEGPLHHYRQOESPSPQOQ--LPPSSQAEGMGQVPRTHRLITLADHC 2056
Dy 181 anqanestrqyegplhhyr-sqgesppqqpplpssqsgmgqvprthrilitladhic 240
Qy 2057 QITQDFARNQVSSQTPOQPTSTFQNSPNSALVSTPVRTKTSNRYSPESQAQSVHQRGP 2116
Dy 241 qitqdfarnqv---psqbststftqspalsaltprtktsrpsqsgstvlhprp 296
Qy 2117 SRVSPENLVKSRGSRGKSPERSHVSEPYEPISPQVPVWVHEKQDILLLSQGAEP 2176
Dy 297 prvspenlvksgsrpgkspershipsepyepispqgppavhekqsdmlslsgrvdp 356
Qy 2177 EQRNDARSPGISYLSYFPTKLENTSPMWKSKQKQEIFRKLNSGGGSDMAAAQPGTEIF 2236
```

|          |                                                                     |                                                                                    |      |
|----------|---------------------------------------------------------------------|------------------------------------------------------------------------------------|------|
| Db       | 357                                                                 | eqrdsrpspsisylpsftkilestpmvkskqelfrklmssgggdsmaaaqpgteif                           | 416  |
| Qy       | 2237                                                                | NLPavTTSGSVSRGHSFADPASNLGLEDIIRKALMGDFDDKVEDHGVVMSQPMGVVPGT                        | 2296 |
| Db       | 417                                                                 | nlpavttsavsrshsfadpasnlgledilrkalmgdfddkvedhgvvmsphvgimpgs                         | 476  |
| Qy       | 2297                                                                | ANTSvVVTSGETRRREGDPSPHSGGVCKPLISKSNRSKSPIPQGGYLGTERPSSVSV                          | 2356 |
| Db       | 477                                                                 | astsvttssearrdegepspha-gvckpklmksnrkskspipqgsylgterpssvsv                          | 535  |
| Qy       | 2357                                                                | HSEGDVHROTPGHAWEDRSGSTGQFPNPLTMRMLSTPTPTACAPSAVNOAAHQH                             | 2416 |
| Db       | 536                                                                 | hsegdvhrctpgvawedrpsstgstqfypnptirmisltstptgtiacapsaitgaahqg                       | 595  |
| Qy       | 2417                                                                | NRIWEREPAPLLSAQYETLSDSD 2440                                                       |      |
| Db       | 596                                                                 | nriwerpapllsaqyetsdsdd 619                                                         |      |
| RESULT   | 5                                                                   |                                                                                    |      |
| AAW18226 |                                                                     |                                                                                    |      |
| ID       | AAW18226                                                            | standard; Protein; 1495 AA.                                                        |      |
| AC       | AAW18226;                                                           |                                                                                    |      |
| XX       | XX                                                                  |                                                                                    |      |
| XX       | XX                                                                  |                                                                                    |      |
| DT       | 24-SEP-1997                                                         | (first entry)                                                                      |      |
| XX       | XX                                                                  |                                                                                    |      |
| DE       | Transcriptional co-repressor SMRT.                                  |                                                                                    |      |
| XX       | XX                                                                  |                                                                                    |      |
| KW       | Silencing mediator for retinoic acid and thyroid hormone receptor;  |                                                                                    |      |
| KW       | SMRT; transcriptional co-repressor.                                 |                                                                                    |      |
| XX       | XX                                                                  |                                                                                    |      |
| OS       | Homo sapiens.                                                       |                                                                                    |      |
| XX       | XX                                                                  |                                                                                    |      |
| FH       | Key                                                                 | Location/Qualifiers                                                                |      |
| FT       | Region                                                              | 1..160                                                                             |      |
| FT       | FT                                                                  | /label= N-terminal_region                                                          |      |
| FT       | FT                                                                  | /note= "proline-rich domain"                                                       |      |
| FT       | Region                                                              | 773..790                                                                           |      |
| FT       | FT                                                                  | /label= ERDR_region                                                                |      |
| FT       | Region                                                              | 812..827                                                                           |      |
| FT       | FT                                                                  | /label= SG_region                                                                  |      |
| FT       | Region                                                              | 1061..1132                                                                         |      |
| FT       | FT                                                                  | /label= glutamine-rich region                                                      |      |
| FT       | Region                                                              | 1201..1495                                                                         |      |
| FT       | FT                                                                  | /label= C-terminal_region                                                          |      |
| FT       | Peptide                                                             | 1330..1376                                                                         |      |
| FT       | FT                                                                  | /note= "alternatively spliced insert not present in the original two-hybrid clone" |      |
| FT       | FT                                                                  |                                                                                    |      |
| FT       | FT                                                                  |                                                                                    |      |
| FT       | FT                                                                  |                                                                                    |      |
| PN       | WO9709418-A1.                                                       |                                                                                    |      |
| XX       | XX                                                                  |                                                                                    |      |
| PD       | 13-MAR-1997.                                                        |                                                                                    |      |
| XX       | XX                                                                  |                                                                                    |      |
| PF       | 24-JUL-1996;                                                        | 96WO-US12371.                                                                      |      |
| XX       | XX                                                                  |                                                                                    |      |
| PR       | 01-SEP-1995;                                                        | 95US-0522726.                                                                      |      |
| XX       | XX                                                                  |                                                                                    |      |
| XX       | (SALK ) SALK INST BIOLOGICAL STUDIES.                               |                                                                                    |      |
| PA       | Chen JD, Evans RM;                                                  |                                                                                    |      |
| PI       | WPI; 1997-192894/17.                                                |                                                                                    |      |
| XX       | XX                                                                  |                                                                                    |      |
| DR       | XX                                                                  |                                                                                    |      |
| XX       | XX                                                                  |                                                                                    |      |
| PT       | New co-suppressor of steroid-thyroid hormone receptor activity -    |                                                                                    |      |
| PT       | also methods for identifying compounds that relieve its suppressant |                                                                                    |      |
| PT       | effect and/or activate receptors                                    |                                                                                    |      |
| XX       | XX                                                                  |                                                                                    |      |
| PS       | Claim 2; Page 40-45; 71pp; English.                                 |                                                                                    |      |
| XX       | XX                                                                  |                                                                                    |      |
| CC       | A novel receptor interacting factor (AAW18226) is designated SMRT.  |                                                                                    |      |





Db 794 kdvdeyrrslvhevgkppqdvddssppskkr-----mhvdfdictkreinyrrsqi 848  
Qy 206 QOOLBEEAAKPP-----EPEKPV-SP-----PPVEOK-----HRsIVQII 239  
Db 849 sedsertsgpsvrhgsfheddpigsprllsvgskpvdkvlpypsnitvreeslkfnp 908  
Qy 240 YDENRKK--AEEAHKIFGLGPKVELPLYNOPSDFKVVHENIK-----TNOVMRK 287  
Db 909 ydsrsreqmadma-----kiklsvlnsedelnrdwsmqkdagfdvdfpsnlikr 959  
Qy 288 KLILFFKRNHARKQREOKICORYDOLMEAKKVDRIENPNRRKAKESKREYVEKQFP 347  
Db 960 dsf-----krsvrdleppevsds-----edgch-----shprasalyessrflldr 1008  
Qy 348 EIRKOREQOERFORVGORGA-----GLSATIARSEHEISEIDGL--SQEN----- 392  
Db 1009 e-dklirderlssalernkfyfaldktitpdkallierakslssreenwsfldwsr 1067  
Qy 393 -----NEKOMRQLSVP-----PMFPAEORRVKFINNGLMEDPMKVVYKDRQFMNVWTOHE 444  
Db 1068 fanfrnnkdkevdapripipswy-mkkkkir--tdsegkmdkdkedheee-----qer 1119  
Qy 445 KEIFKDKFIQHPKNGLLIASYLERSVPCVLYLYLTTKNENYKALVRNKGKRRGRNQ 504  
Db 1120 qefasfih-----ssifeg-----dskrlqlerkeedsdfisrlygkqtsagan 1167  
Qy 505 IARPSQEEKVEEKEDKAEKTEKBEKDEEKEDEKEDSKENTKEKDKIDGTABETEER 564  
Db 1168 sttdsiqepvvlifshfmeht-rmqkkkkkdkqpkvekedtenhpktpesapenks 1226  
Qy 565 EQATPRG-----RKYANSQGRKRGITRSMWNEAASAAAAAATEE 606  
Db 1227 elktppsgpssvvtvlesapsalektgtktveaplvteektvpatvseeakpasep 1286  
Qy 607 pppp-----Lpp----- 613  
Db 1287 apavpqlqevdlppgadpkeaaampagveegsgdppylakpptpgasfsqaesnv 1346  
Qy 614 PPEPTSTPEVETSRWTEEMEYAK-----KGLVEHGRNAAIA----- 651  
Db 1347 dpepdstqlspakpakeeaneapkaekpadataadaepdanqkaeaapesqppasedlevdp 1406  
Qy 652 -----KMWGTSKSAQCKNFYFNKRRHNDLNLQHKQKTSRPRERD-----v 696  
Db 1407 pvaakdkkpnkskrtppvqaaavisekvpvtrkseridreklrksnspgreaqklllek 1466  
Qy 697 SQCESVASTVSAQEDIEDIE----- 715  
Db 1467 meakitrtasknaadliehpelplstrrrrvrsyvatmgdhenrsrvkpevqprv 1526  
Qy 716 -----ASNENBENPEDSEVAVKPSED--SPENATS 743  
Db 1527 trklrelqelaaavtprgrpbktrrradeeeneakepaetlkppegvrsprskt 1586  
Qy 744 -----RGNTPEAVEL---EPTTETAPTSTSLAVPSTKPAEDSEVETQVNDISIAE 791  
Db 1587 aagggpggkkgnekpvkdrteatpveatvvgpqlgvkessmepkaaeaeagseqkdrkdag 1646  
Qy 792 TAEQMDVQDQEHSAEESGVCPDPATKADSDVDEVVRVPENHASKVEGDNTKER----- 844  
Db 1647 tdkn-----ppetapvevvekkpapekkskgrsrnsrlavdksa 1688  
Qy 845 DLDRASEKVEPRD-----EDLVAAQOINAQRPE--PQSDNDSATCSADE-DVDGPE 894  
Db 1689 slknvdaavprgaaqaqageresgvva--vspekseapkedgissqlksdpvdpdkpe 1746  
Qy 895 RORFPMDKPSL-----LNFTGSILVSSP--LKPNPLDLPOLQ 931  
Db 1747 kedvsasopspeatqlakmeleqavehiaklaeasasaaykadapeglapedrdkphq 1806  
Qy 932 -----HRAAVIPPM---VSCTCPNIPIGTPVSGYALYQRIKAMHESALLEEORQOEID 984  
Db 1807 asetelaaagsiindisgepenfapppypyg-----esqtd 1843

Qy 985 LECRSSTSPCGTSKSPNREWEVLQAPHOLITNLPEGVRLPT---TRTPRP---PPLIP 1038  
Db 1844 lq-----ppaga-----qalqpseegmetdeavsgileateassrppvnapd--p 1888  
Qy 1039 SKTTVASEKPSFIMGSIQSQTPTGYLTSHN--OASVTOETPKPSVSGISIGLPRQOES 1096  
Db 1889 sagptdtke-----argnsse-----tshsvpeakgsk-----vevllvrkdkg 1928  
Qy 1097 AKSATLPYIKO-----EEFSPRSONSOPEGLLVYRAQHEGVVVRGTAGAIQE---GS 1143  
Db 1929 rqtctrsrkntnkvvapveshpesnaqges---paanegtvtvqheapegkeqse 1985  
Qy 1144 ITRGTP-----TSKIVESIPSLRGSITQCTPALPOTGPTALVRKGSISRMPIEDSSPEK 1199  
Db 1986 kphstppgsctsdls--kipstens--sqeisveert--ptkasvpdpdp--pppapavd 2038  
Qy 1200 GREEAASGHVYIEG-----KSGHILSYDNINKNAREGTRSPRTAHEISLKRSYVESVEGNI 1254  
Db 2039 eepqarfvrhsliesdpvtppsdpisipiptlpsvtaaklspvva---sggiphqspptkv 2095  
Qy 1255 KOGMSMRSP--VSAPLEGLICRALPRGSPHDLKE-----RTVLSSGSIQGTTPRATTES 1307  
Db 2096 tewitrcpgepraqscp-----spalppdtkasdvdstsrlrkilmdpkysatsavtats 2150  
Qy 1308 FEDGLYKPKQIK---RESPPIRAFEGAITKGPYDGITTIKEMGRSIIHEIPRQDILTOES 1364  
Db 2151 vttaiaepysaapclheapp-----ppvdskkpleekctappvttnns--eigasevivaad 2203  
Qy 1365 RKTPTVWOSTPRIEGSISQGTPIKFD--NNSQSAIKINVKSLITGPGSKLSRGMPPLEIV 1423  
Db 2204 ke--kvapviapkitsvisr--mpvsiidlensqkiklappatqltglvsaltglvnsvl 2260  
Qy 1424 PEN-LKVVERGKYEDVKAGETVRSRHTSVWS--SGP-SVLRSTLHEAPKAQLSPGIYDDT 1479  
Db 2261 pvnalkpvgksgvttlk-----slvtpagpvnvllkpgvnav-----ltgvpv 2301  
Qy 1480 SARRTPV-----SYQNTMSRSGSPMMNRT--SDVTIPPNKSTNHERKSTLTP----- 1523  
Db 2302 nvlttpevnatvgtvnaaagptvnaaasavnatasavtvtagavtaasgvgvttgtvtmag 2361  
Qy 1524 -----TQRESIPAKSPV-PGVDPVVSHPFPPHHRGSTA-----GEVYWSHL- 1564  
Db 2362 aviapstkckgrasanensrfhpgsmpvlddrpad--agsgaglrvtseggvllvsyg 2418  
Qy 1565 -----PTOLDPAMPPHRALDPAAYL-FOROLS-----PTPGYPSOYQLY 1604  
Db 2419 qktegpqrisaki---sqippasamdiefgqsvksqvkpdsvtasqppskg--paapagy 2474  
Qy 1605 AMENTROTILNDYITTSQOMQVNLDPDVARGLSPR---EQP-----LGLPYPATRG-----I 1652  
Db 2475 anvathstlv---ltaq--tynaspvisvskadrpslekpepihlsvstptvggtkvvl 2529  
Qy 1653 IDLNMWPT-----ILVPHPGGISTPPMDRITYITPGTITFPPRPYN--SASMSGHPH 1705  
Db 2530 tggintppvlvhnqilvtlpsivttkkladpvtliketkvl---qpanlgstltphhpa 2586  
Qy 1706 LAAASASEREREREK--ERERERIAAASDDLVRPGSEOPGRPGSHGVYRSPSPSVR 1762  
Db 2587 lpsklptevnhvpsgpsipadrtvshlaaakldah-----sprpspgpspsfrashps-s 2641  
Qy 1763 TQETMLQORPSVF---QGTNGTSVITPLDPTAQLRIMPLPAGGPSISQGLPAS-----R 1813  
Db 2642 tastalstnatvmlaaglpvqqfissihp-eqsvlmp---phsitqvtvshlsqgevr 2696  
Qy 1814 YNTAADALAALVDAASAPQMDVSKTESKHEAARLEENLRSSAAVSEQOOLEOKTLEV 1873  
Db 2697 mnt-----ptlpsityslrpealh-----spraplpqpg-----i 2726  
Qy 1874 EKRSVQCLYTTSSAPSGKPPQPHSSVVYSEAGKDKGPPPKSYEEELRTRGTTTTAANFI 1933  
Db 2727 evrapqrastpqpapagpv-----alasqhpp-----eeevhyhlparatapvq 2771

QY 1934 DVIIITQIASDKARERGSQSSSSLSHRYETPDAIEVISP-----ASSP----- 1982  
Db 2772 sevlvmq-----seylhpytprdrvimhphvtavseqpraadgv 2813  
QY 1983 --APPOEKLQTVQ--EVVKANQAEENDPTR-----QYEGPLHHYRPOQESPQOQLPPS 2033  
Db 2814 kvppask-apqpgkaektadkaaptptpavpvpvpipapapaphgearililtvps 2872  
QY 2034 SOAEGMGVPRTHRLITLADHICQII-----TQDFARNQVSSQTPOQPTSTFQNSPSA 2087  
Db 2873 nqlqgipitpp-----vvvthgqvihssgelqfgyrygdirtyhpbpaqltht-qfpaas 2926  
QY 2088 LVSTPVRTKTSNR-----YSPESQAQSVHHORPGRSVSPENIVDRSGSRPG----- 2134  
Db 2927 svglpsrtkaagpppgeplqpqpqvstcpaqpapcpqpsql-----gqpqppss 2980  
QY 2135 KSPERSHVSSEYEPISPPQVPVHKEODSLLLSQGAEPAEQRNDAR-----SP 2185  
Db 2981 kmpqvsqaekgtqtgveqprlpagp-----anrppehtqvraqaetgptsfppsp 3031  
QY 2186 GSISYLPSEFFTKLENTSPWVSKKQEIFRKLNSGSGSDMAAQPCTEINLPAVTTSG 2245  
Db 3032 vsvsmkpdlpvsl-----ptqapkpqlf--vpttsgpstpgqlvlphtefgpaqkdssp 3085  
QY 2246 SVSSR 2250  
Db 3086 hltsq 3090

RESULT 7  
ID AAB35408  
XX AAB35408 standard; Protein; 2819 AA.  
AC AAB35408;  
DT 23-MAY-2001 (first entry)  
DE Human 07CG27 gene protein.  
KW Human; 07CG27 gene; chromosome 1; HPC1 region; prostate cancer;  
KW oncogene.  
OS Homo sapiens.  
XX  
XX  
XX WO200116291-A2.  
XX  
XX PD 08-MAR-2001.  
XX  
XX 25-AUG-2000; 2000MO-US23291.  
XX  
XX PR 27-AUG-1999; 99US-0151049.  
XX  
XX PA (MYRI-) MYRIAD GENETICS INC.  
XX (HOSP-) HOSPITAL FOR SICK CHILDREN.  
XX  
XX Tavtiglian SV, Swedlund B, Simard J, Rommens JM;  
XX WPI; 2001-226682/23.  
XX N-PSDB; RAF28060.  
XX  
XX Novel human prostate cancer marker gene termed as 07CG27 gene, useful  
XX for screening mutations in the gene in diagnosis of a predisposition to  
XX cancer -  
XX  
XX Claim 1; Page 91-99; 99pp; English.  
XX  
XX The present invention provides the protein and coding sequences of the  
XX human 07CG27 oncogene. This gene is found at the HPC1 region of  
XX chromosome 1. The sequences can be used in the diagnosis and  
XX identification of treatments for prostate cancer. The present sequence is  
XX the 07CG27 protein.  
XX  
XX Sequence 2819 AA;

Query Match 4.0%; Score 507.5; DB 22; Length 2819;  
Best Local Similarity 18.5%; Pred. No. 4.2e-19;  
Matches 557; Conservative 350; Mismatches 962; Indels 1147; Gaps 135;

QY 6 YPNOGAFSTQSRVPHSHVQVTFPNTRHQ--QEFVAVPDYRSSHLEVSQASQL----- 56  
Db 260 ypllhgm-----rfpp-sisetnkglrgrpppswaseperpsllsaseikdkfdnl 313  
QY 57 -----LQOQQOQLRRRPSLSSEFHPGSDRQOERTSYEPFHPGSPVDHDSLESKR 108  
Db 314 daeadeqwagmevdyteqinfdsdddeggnspkennse-----dqgskasen 362  
QY 109 PRLQOVSDSHFORVSAAVLPL---VHPLPEGLRASADAKKDPAP---GGHEAPSSP--- 159  
Db 363 nenketdevantksssqipagpavakvpyg-----kgpsfnqertgtshtpppkl 414  
QY 160 ISGQPCGDDQNASPSK---LSKEELIQSDRVDREIAKVEQOILKKKKQOQ---LEE 211  
Db 415 laqhhppdrqavgrpgpfpskqvad-----edeiwwkrrrqqgseisaave 462  
QY 212 EAAKPPEPEKVPSPVPPVQKHRSIVQIITYDENRKK--AEAAHKIFEGGLGPKVPLYNQ 269  
Db 463 rarkrre-----eeerr-----meeqrkaacaekikrideklg-----ilekqp 501  
QY 270 SDTKVYHENIKTNQVMRKLLILFPKRRNHARKQREOKICQRYDQLMEAEKKVDRIENNP 329  
Db 502 speel-----rerereker-----ekelekeqe 528  
QY 330 RRAKESKTREYVEKQFPEIRKQEQERFORVGAGLSATIAARSEHSEIIDGSE 389  
Db 529 rekere-kdrerqekekekekeqremek-----erkqekeke-----ler 571  
QY 390 QENNEKOMRQLSVIPPMFMFADAEQRRVKFINNGLMEDPMKYKDRQFMVWVTDHEKEIFK 449  
Db 572 qkekekelqkmk-----equekecel 591  
QY 450 DKFTQHPKNFGLIASYLERKSVPCVLYLYLTKKNNYKALVRRNYGKRRGNQOIARPS 509  
Db 592 ek-----erekl----- 598  
QY 510 QEEKVEEKEDAKTEKKEE---KKDEEKDEKEDSKENTKEKDKIDGTGTAETEEREQ 566  
Db 599 -eekieprepnlepmvkekeseensceneepvfrqdsnrsekeatpv---vhetepesg 654  
QY 567 ATPRG-----RKANSQGRKGRITRSMTNAAASAAAA 601  
Db 655 sqprpavlsgyfkqfkslpprfqrqeqmkgqgqqgqgqvlpqtvpssstsv--- 711  
QY 602 AATEEPPPPPLPPEPISTEPVETS-----RWTEEMEVAKKGLVEHGRNWAIAKMWG 655  
Db 712 -----pppphrplyqmpqhphqhlasmgfdprwlmmsymvpr---mmsgrpamdippi-- 762  
QY 656 TKSEAQCKNFYFNKRRHNLNLLQHKQKTSRKPREERDVSOCESVASTVSAQDEDEIE 715  
Db 763 -----hpgmipppklmrid-----qmegspnssesfehia 792  
QY 716 ASNEENPEDSEVAVKPESDPENATSRGNTPEVALEP---TTETAPSTSPSLAVPST 772  
Db 793 rsardhaaisleprmlwgsdpyphaepqattpkate-epedvrsaaldegqitaaysv 851  
QY 773 K-----PAED-----ESVETQVNDNSISAETAEQMDVQOQHSAAEGSVCDPPATKADSV 822  
Db 852 ehnlqleahpkadfirreseaqvqlsrsv---dvprhhtdannqsacfeapqktlisa 908  
QY 823 DVEVRVFNHASKVEGDNTERKDRASEKVEPRDEDLVVAQGINAORPEPQSDNS--SA 881  
Db 909 pqeeri-----savesqpskrsvshgshntkqkpe-----qrsepsagipkvt 953  
QY 882 TCSADEVDGEPERQRPFPMDSKPSLLNPTGSLVSSPLKPNPLDLPOLOHRAAVIPPM- 940  
Db 954 rcldskepierpe-----ekp---kkegfirsgepkpekvyksksetrwprrpsn 1002





[illegible]

|          |             |                                                                      |      |
|----------|-------------|----------------------------------------------------------------------|------|
| Db       | 2758        | gedgikalkesakqtppaasvt--ysr---rrprapresaqledlagfkdpaghtees           | 2814 |
| Qy       | 2165        | LLLLSORGAEPAEORNDARSPGISYLPSPFFTKLENTS-----PMVSKKOEIFRKLNS           | 2218 |
| Db       | 2815        | m-----tdtktkipkss-----peledatsskrrprtraqkvevkeella                   | 2857 |
| Qy       | 2219        | SG-----GGSDMAAQP-----GTIFNLPA---VTTSGSVSSRGHSFADPASNGLIED            | 2265 |
| Db       | 2858        | vgkltqtsgetthtdkepvggkgtkafqkapkrnvdaedvigstrqprapkekaqpld           | 2917 |
| Qy       | 2266        | IIRKALMGSPDDKVEDHGVMVQPMGVPECTAN---TSVVTSGETRREEGDPSPHSGGVC          | 2322 |
| Db       | 2918        | -----lasfqe-----lsqtphgteelangaadstfapktqpsgkplksrrvl                | 2963 |
| Qy       | 2323        | K-PKLISKSNRSKSPKIPGGYLGTERP                                          | 2350 |
| Db       | 2964        | rapkvepgdvvstrdpvkqsksntslp                                          | 2992 |
| RESULT 9 |             |                                                                      |      |
| ID       | AAR53557    | standard; protein; 73 AA.                                            |      |
| XX       | XX          |                                                                      |      |
| AC       | AAR53557;   |                                                                      |      |
| XX       | XX          |                                                                      |      |
| DT       | 10-FEB-1995 | (first entry)                                                        |      |
| XX       | XX          |                                                                      |      |
| DE       | XX          |                                                                      |      |
| XX       | XX          | Thyroid hormone receptor-interacting protein.                        |      |
| XX       | XX          | nuclear thyroid hormone interacting proteins; TR; JLI; JLI2;         |      |
| KW       | XX          | transcriptional coactivator; treatment; diagnosis;                   |      |
| KW       | XX          | thyroid related disorders; modulation; thyroid hormone receptor;     |      |
| XX       | XX          | nuclear hormone receptor; isolation.                                 |      |
| OS       | XX          | Chimeric Homo sapiens.                                               |      |
| OS       | XX          | Chimeric Bacterial sp.                                               |      |
| XX       | XX          | W09410338-A.                                                         |      |
| XX       | XX          | 11-MAY-1994.                                                         |      |
| PD       | XX          |                                                                      |      |
| PF       | XX          | 29-OCT-1993; 93WO-US10443.                                           |      |
| XX       | XX          |                                                                      |      |
| PR       | XX          | 30-OCT-1992; 92US-0969136.                                           |      |
| XX       | XX          |                                                                      |      |
| PA       | XX          | (GEO ) GEN HOSPITAL CORP.                                            |      |
| PI       | XX          | Lee JW, Moore DD;                                                    |      |
| XX       | XX          | WPI; 1994-199808/24.                                                 |      |
| DR       | XX          | N-PSDB; AAQ63705.                                                    |      |
| DR       | XX          |                                                                      |      |
| XX       | XX          |                                                                      |      |
| PT       | XX          | Nuclear hormone receptor interacting polypeptides, esp. thyroid      |      |
| PT       | XX          | hormone-interacting proteins (TRs) - for identifying proteins        |      |
| PT       | XX          | useful in treatment and diagnosis of thyroid related disorders by    |      |
| PT       | XX          | inoculating thyroid hormone receptor activity                        |      |
| XX       | XX          |                                                                      |      |
| PS       | XX          | Claim 35; Page 61; 105pp; English.                                   |      |
| XX       | XX          |                                                                      |      |
| CC       | XX          | This sequence shows the partial amino acid sequence of a thyroid     |      |
| CC       | XX          | hormone (TR) interacting protein. TR-interacting proteins physically |      |
| CC       | XX          | associate with thyroid hormone receptor. Nearly all the fusion cDNAs |      |
| CC       | XX          | showed very strong dependence on hormone activation. The proteins    |      |
| CC       | XX          | can be used in an in vivo trap system for the isolation of proteins  |      |
| CC       | XX          | which associate with any nuclear hormone receptor. The proteins and  |      |
| CC       | XX          | Abs may be used to treat or diagnose thyroid disorders; and to       |      |
| CC       | XX          | modulate thyroid hormone receptor activity.                          |      |
| XX       | XX          | Sequence 73 AA;                                                      |      |
| SQ       | XX          |                                                                      |      |

Query Match 2.9%; Score 370; DB 15; Length 73;  
Best Local Similarity 100.0%; Pred. No. 1.2e-13;

Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2231 PGTEIFNLPAVTTSGVSSRGHSFADPASNLGLIEDIIRKALMGSDFKVEDHGVVMSQPM 2290  
Db 1 pgtelfnlpavttsgvssrghsfadvpsnlgledliirkalmgsfddkvedhgvvmsqpm 60  
|||||

QY 2291 GVVPGTANTSVVT 2303  
Db 61 gvvpgtantsvvt 73  
|||||

RESULT 10  
AAW40596  
ID AAY40596 standard; protein; 73 AA.

XX AC AAY40596;

XX DT 03-DEC-1999 (first entry)

XX DE Partial amino acid sequence of TR-interacting protein S243b.

XX KW JLI1 protein; thyroid hormone receptor; trap assay; therapeutic;  
KW nuclear hormone receptor protein; TR-interacting protein; human;  
KW thyroid disorder.

XX OS Homo sapiens.

XX PN US5962256-A.

XX PR 05-OCT-1999.

XX PF 06-JUN-1995; 95US-0471613.

XX PR 04-APR-1994; 94US-0222719.

XX PR 30-OCT-1992; 92US-0969136.

XX PA (GEHO ) GEN HOSPITAL CORP.

XX PI Lee JW, Moore DD;

XX DR WPI; 1999-571268/48.

XX DR N-PSDB; AAZ07582.

XX PT Purified DNA comprising a sequence encoding a protein which  
PT specifically interacts with a thyroid hormone receptor -

XX PS Disclosure; Fig 28; 68pp; English.

XX CC The invention provides a JLI1 protein which specifically interacts with a  
CC thyroid hormone receptor in an in vivo trap assay. The JLI1 protein can  
CC be recombinantly produced by standard recombinant methodology. The JLI1  
CC protein is used in a trap assay for determining whether a test protein  
CC is capable of interacting with a nuclear hormone receptor protein, and  
CC may also be used as a therapeutic peptide for treating thyroid  
CC disorders. Sequences AAY40572-596 represent partial amino acid sequences  
CC of thyroid hormone receptor (TR)-interacting proteins.

XX SQ Sequence 73 AA;

Query Match 2.9%; Score 370; DB 20; Length 73;  
Best Local Similarity 100.0%; Pred. No. 1.2e-13;  
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2231 PGTEIFNLPAVTTSGVSSRGHSFADPASNLGLIEDIIRKALMGSDFKVEDHGVVMSQPM 2290  
Db 1 pgtelfnlpavttsgvssrghsfadvpsnlgledliirkalmgsfddkvedhgvvmsqpm 60  
|||||

QY 2291 GVVPGTANTSVVT 2303  
Db 61 gvvpgtantsvvt 73  
|||||

RESULT 11  
AAW92403

XX ID AAW92403 standard; Protein; 73 AA.

XX AC AAW92403;

XX DT 21-APR-1999 (first entry)

XX DE Human TR-interacting protein S243b.

XX KW Thyroid hormone receptor-interacting protein; TR-interacting protein;  
KW JLI1; human; interaction trap assay; treatment; thyroid disorder; S243b;  
KW hyperthyroidism.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
FT Protein 1..73

FT FT /note= "Partial coding sequence, no start or stop  
FT FT codons given"

XX PN US5866686-A.

XX PD 02-FEB-1999.

XX PF 06-JUN-1995; 95US-0470925.

XX PR 04-APR-1994; 94US-0222719.

XX PR 30-OCT-1992; 92US-0969136.

XX PR 06-JUN-1995; 95US-0470925.

XX PA (GEHO ) GEN HOSPITAL CORP.

XX PI Lee JW, Moore DD;

XX DR WPI; 1999-142013/12.

XX DR N-PSDB; AAX01894.

XX PT JLI1 protein - that interacts with thyroid hormone receptor

XX PS Disclosure; Fig 28; 67pp; English.

XX CC This sequence represents a human thyroid hormone receptor (TR-interacting  
CC protein) which is found to interact with a novel human JLI1 protein in an  
CC in vivo interaction trap assay. The JLI1 protein is potentially useful for  
CC the treatment of hyperthyroidism or thyroid disorders.

XX SQ Sequence 73 AA;

Query Match 2.9%; Score 370; DB 20; Length 73;  
Best Local Similarity 100.0%; Pred. No. 1.2e-13;  
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2231 PGTEIFNLPAVTTSGVSSRGHSFADPASNLGLIEDIIRKALMGSDFKVEDHGVVMSQPM 2290  
Db 1 pgtelfnlpavttsgvssrghsfadvpsnlgledliirkalmgsfddkvedhgvvmsqpm 60  
|||||

QY 2291 GVVPGTANTSVVT 2303  
Db 61 gvvpgtantsvvt 73  
|||||

RESULT 12

XX ID AAW85115 standard; Protein; 73 AA.

XX AC AAW85115;

XX DT 09-FEB-1999 (first entry)

XX DE Thyroid hormone receptor-interacting protein S243b partial sequence.  
XX

KW Thyroid hormone receptor-interacting protein; S243b;  
 KW nuclear hormone receptor protein; screen.  
 OS Homo sapiens.

XX US5846711-A.  
 XX 08-DEC-1998.

XX 04-APR-1994; 94US-0222719.  
 XX 04-APR-1994; 94US-0222719.  
 PR 30-OCT-1992; 92US-0969136.

XX (GEO) GEN HOSPITAL CORP.

XX Lee JW, Moore DD;  
 XX WPI; 1999-059040/05.

DR N-PSDB; AAV82605.  
 XX Screening assay for nuclear hormone receptor modulators - using  
 XX cells containing reporter gene construct

XX Disclosure; Fig 28; 69pp; English.  
 XX The present sequence represents thyroid hormone receptor-interacting  
 XX protein S243b. The protein was identified using the method of the  
 XX invention. The method is used to determine if a test protein is capable  
 XX of interacting with a nuclear hormone receptor protein in a  
 XX ligand-dependent manner. The method is used especially to screen for  
 XX proteins that interact with thyroid hormone receptors in a  
 XX ligand-dependent or ligand-sensitive manner.

XX Sequence 73 AA;

Query Match 2.9%; Score 370; DB 20; Length 73;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-13;  
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2231 PCTEFLNPAVTTSGVSSRGHSFADPASNLGLIEDIIRKALMGSDFKVEDHGVVMSQPM 2290  
 Db 1 pgtelnlpavttsgvssrghsfadpasnlgledliirkalmgsfdkvedhgvvmsqpm 60  
 QY 2291 GVPPTANTSVVT 2303  
 Db 61 gvpptantsvvt 73

RESULT 13  
 AAR26052  
 ID AAR26052 standard; Protein; 2843 AA.

XX AAR26052;  
 XX 28-JAN-1993 (first entry)

XX APC gene product in familial adenomatous polyposis.  
 XX neoplasm; cancer; oncogene; tumour; growth; detection; diagnosis;  
 KW prognosis; treatment; sporadic colorectal carcinomas; ss.

XX Homo sapiens.  
 XX WO9213103-A.

XX 06-AUG-1992.

XX 16-JAN-1992; 92WO-US00376.

XX 16-JAN-1991; 91GB-0000963.

PR 08-AUG-1991; 91US-0741940.

XX (CANC-) CANCER INST.  
 PA (ICIL) IMPERIAL CHEM IND PLC.  
 PA (UJO) UNIV JOHNS HOPKINS.  
 PA (UTAH) UNIV UTAH.

XX Albertsen H, Anand R, Carlson ML, Groden JL, Hedge PJ;  
 PI Joslyng, Kinzler KW, Markham A, Nakamura Y, Thliveris A;  
 PI Vogelstein B, Whitehl, Markham AF;  
 XX WPI; 1992-284685/34.

DR N-PSDB; AAQ27234.  
 XX Detection of somatic and germ-line alterations of human APC gene  
 XX - used to diagnose, treat and study familial adenomatous  
 XX polyposis and sporadic colorectal cancer

XX Disclosure; Page 47; 132pp; English.

XX This sequence is encoded by the APC (Adenomatous Polyposis Coli)  
 XX gene associated with tumorigenesis, found on chromosome 5q.  
 XX The sequence may be mutated by deletions insertions, inversions, or  
 XX point mutations of the gene. The APC gene is expressed in most normal  
 XX tissues as well suggesting that APC is a tumour suppressor.

XX Sequence 2843 AA;

Query Match 2.9%; Score 368; DB 13; Length 2843;  
 Best Local Similarity 19.4%; Pred. No. 1.8e-11;  
 Matches 418; Conservative 281; Mismatches 797; Indels 664; Gaps 98;

QY 347 PEIRKQREOE-RFORVGORGAGLSATIARSEH-----EISEIIDGLSEQENNEK 395  
 Db 981 psiesyeddeskfcsygyqypadlahkhsanhdndgdeldtptnlsykysdedsgr 1040  
 QY 396 QMRQLSVIPMMFDAQRRVKFINNGLMEDPMKYKDRQFMNVTDHE--KEIFKDFEI 453  
 Db 1041 q-----spsqnerwarphkhiiedelkqseqgrngsttvpvtestdkhl 1087  
 QY 454 QHPKNFGLIASYLERKSPDCVLYLLTKNENKALVRRNYKRRGNQOIAEPSQEK 513  
 Db 1088 kfqphfg-----qqecvspy-----rsrgangsetnrvsnhngnqvsglqde- 1132  
 QY 514 VEEKEDEKAETK--BEEKDEEKEDEKEDSKENTKEKIDGTABETEEREQATPRG 571  
 Db 1133 -dyeddkptnyseryseeqheeeerptnysikyn-eekrhvdpidyslkaytdipss 1190  
 QY 572 R-----KTANSQRRKGRITRSMNEAAS-----AAAAATEEPPPLPPP 614  
 Db 1191 qkqsfkskssgsktehmssssentstpsnakrnqlhpsaqsrgsqgpkkaatck 1250  
 QY 615 PEPISTEPVEV-----SRWT-----EEMEVAKKGLVEHGRNWAATKM--- 653  
 Db 1251 vssinqetigtvcvtedtpicfscsslsssaedeigcndtqeadsantlqlaekgk 1310  
 QY 654 VGTKSE-----AQCKNFYFNKRRHNLNLOQHK-----QKTSR 688  
 Db 1311 igtrsaedpysevpavsqhprtkssrlqgsslsesarhkavfsgakspksaqtpk 1370  
 QY 689 KPREE-----RDVQSCEVASTVSAQDEDEIEASNEENPEDEVEAVKPE--DSPEN 740  
 Db 1371 sppehyvqetplmfrctsv-ssidsfesrsiasvqpcsgmvsiglisldpdpqg 1429  
 QY 741 A--TSRGNTPEAVELEPTTETAPSTPSLAVPTKPADESVETQVNDISIETAEQMDV 798  
 Db 1430 tmppsrsktppp---ppqatqkrevpknkaptakresgpkgaavnaavg-rvqvlpda 1485  
 QY 799 DQOEHSAEE-----GSCV-----DPPPATKADSVDEVR---VPEN-HASKVEGDN 840  
 Db 1486 dtllhfatestpdpfscsslsalsldpfiqk----dveirmpvqendngneteseq 1541  
 QY 841 TKE--RDLDRASEKVEPRDEDLVVAQINAOQRPFPQSDNDSSATCSADEVD----- 890

```
Db 1542 pkesnenqekaektidsekdl-----ddsd-----dddieileeciis 1581
Qy 891 -----GPERQRMF-----PMDSKPSLLNPTGSIILVSSPLKPNPLDLQLOHRAVI 937
Db 1582 amptksrkgkpaqtaasklpppvarkpsql-pvykllpsq-----nrl-----qpqkhsft 1633
Qy 938 P-----PMVSC-----TPCNPICPTPVSGYAL-----YQRHKAHME 969
Db 1634 pgddmrvycvegtpinfstatisldtiesppnelaagevrgagsggefekrdtpte 1693
Qy 970 SAL-----LEEQRQRQBIDLECRSSTSPCGTSKSPNREWEVL----- 1007
Db 1694 grstdeaggtssvtipelddnkaeegdilaecinsampgkshkpfirvkkimdvqqa 1753
Qy 1008 ---OPAPHQLITNLPEGVRLTTRTPPPPLI-----P 1038
Db 1754 sassaapnk-----nqldgkkkxtpvkipqnteytrvrknadsknnlnaerfsdnkd 1810
Qy 1039 SSKTTVASEKPSF-----IMGGIS-----QCTPGTYLTSHNOASVTOETP 1079
Db 1811 skknlknnskdndklpnndrvrgsfafdsphhytpiegt-p-ycfsrnd----- 1860
Qy 1080 KPSVGSISLGLPQOESAKSATLPYIKOEFSRQNSQPEGLLVRAQHEGVVGTAGAI 1139
Db 1861 ---sissldfddvdlrskaelrkakenkesakvtshte--ltsnqgsa---nktqai 1913
Qy 1140 QEGSITRGTPTSKISVESIPSLRSITOGTPTALQGTPTALVK-----G 1185
Db 1914 akqpinrgqpkpiliqkgs-----tfqsskdipdrgaatdekqnfaintpvcfshns 1967
Qy 1186 SISRM-----PIEDSPEKREAAKSGHYIYEGKSGHI----- 1219
Db 1968 slsslsiddqennnenepiketepdsggepskpaqsgaypaksfhvedtpvcfsrns 2027
Qy 1220 ---LSYDN-----IKNAGRTSRPTAHEISLKRYSVESVEGNIKQMSRSPVSAPL 1269
Db 2028 sslsidseddllqecissampkksrkkgdnekhprnmngilgedltldldkdiqrd 2087
Qy 1270 EGLICRALPRGSHDLKERTVLSGSIWGTPTPRATTESFEDGLKYPKQIKRESPPIRAFE 1329
Db 2088 se---hglspdsenfdkalgansivsslhqaaaaac-----lsrqassdsdlsisk 2139
Qy 1330 GAITKGKPYDGIITTIKMGRSIHEIPRODILTQESRKTPPEVQSTRIEIGSISQGTPIK 1389
Db 2140 sgislgspf-----hltpdeekpftsnkqprilk-----pgekstletkk 2180
Qy 1390 FDNNS-GQSAIKHNKVSILITCP-----SKLSRGM-PPEIIVPENIKVVERGYEDVKAGET 1443
Db 2181 ieseskgikggkvyksliitkvrsnseisgmkpqlq---anmpsirsrg----- 2227
Qy 1444 VRSRHTSVVSSGSPVLRSLTLHEAPKAQLSPGIIYDDTSARRTPVSVQNTMRSGSPMNRTS 1503
Db 2228 -----rtmih-----ipgv-----rnsststspvskkqp 2251
Qy 1504 DVTIPPNNKSTNHERKSTLTPT-QRESIPAK-SPVPGVDVSVVSHSPFDPHHRGSTAGEYVW 1561
Db 2252 plktpaskspsegqtattspgkpsvkselspvartqtsqigsskapsrsgr----- 2305
Qy 1562 SHLPTQDLPAMPFHRALDPAAAYLFQRLSPTRGYSQYOLYAMENTRQTILNDYITSQ 1621
Db 2306 -----dstpsrpaqql-----sr 2319
Qy 1622 QMQVNLRPDVA---RGLSPREQPLGLPYPATRGIIIDLTNMPPTILVPHPGGTSTPPMDRI 1678
Db 2320 piqpggrnsgispgngisppnk-----lslqprt-sspsstaskssggskm 2364
Qy 1679 TYIPGTQITFPFRPNYSASMSGPHTHLAAASAERERERE-----KERER 1726
Db 2365 sy-----tspggrqmsqqlt--kqtglsknassiprsesaskglnqnmngnngankkvel 2416
Qy 1727 ERIAAASDDLRLRPGCEQPRGSHGYSVSPSPSVRQVQTMLOQRPSVFQGTNGTSTVTP 1786
```

```
Db 2417 srmstks-----gsesd-----rserpvlvrqstfikeaps----- 2449
Qy 1787 LDPTAQLRIMPLPAGGSIISQGLPASRYNTAADAALAAALVDAASAPQMDVSKTESKHEA 1846
Db 2450 ---pclr-rkieesasfesis---psrpsasptrsqatpvlspdpmsis-thssvqag 2502
Qy 1847 A---RLEENLRSAVSEQQOLEQTLVEKRSVQCLYTSSAFFSPKQPQHSSVYVSEAG 1904
Db 2503 gwrklppnl---sptieyndgrpakrhdiar-----shsesprlipnrsrgtwkrehs 2552
Qy 1905 KDKGPPKPSRYEELRGRKTTITAAAN-----FIDVIITROIADKADAREGSSQSSD 1956
Db 2553 khssslp--rvstwrtrtgssssillsassesekaksekdekhvnsisgtkgskenqvsaag 2610
Qy 1957 SSSSLSSHRYE-TPSDALEVISPASSAPPQOEKLTQYQPEVVVKANQAEENDTRQYEGPLH 2015
Db 2611 twrkikeneisfptsntsqtsvsgatngaesktllyqmapavsk---tedvvrriedcpin 2667
Qy 2016 HYRPOQESPSPOQQLPP-----SSQAEGMGQVPRTHRLITLADHI 2055
Db 2668 npr---sgrsptgntppvidsvsekanpnikdskdnqakqnvngsvpm--rtvglenrl 2722
Qy 2056 CQITQDFARNQVSSQTPQPPPT--STFQNSPSALVST---PVRTKTSNRVSPESQASQSV 2110
Db 2723 tsfiqvd-----apdqkgteikpggnupvpvsetnespivert-----pfssssss 2768
Qy 2111 HHQRP---GSRVSPENLVDKSRGSRPKSPERSHVSSEPVPEPISPPQVPPVH--EKQDS 2164
Db 2769 khspsgtvaarvcfnfy-----npsrksadsatsarpsqip-tpvnnntkkrrds 2818

RESULT 14
AAW35392
ID AAW35392 standard; Protein; 2843 AA.
XX
AC AAW35392;
XX
DT 11-MAR-1998 (first entry)
XX
DE Human adenomatous Polyposis coli gene product.
XX
KW Human; adenomatous Polyposis coli; APC; diagnosis; prognosis;
KW neoplastic tissue; tumour tissue; tumour repressor; mutation;
KW sporadic colorectal cancer; detection.
XX
OS Homo sapiens.
XX
PN US5648212-A.
XX
PD 15-JUL-1997.
XX
PF 12-AUG-1994; 94US-0289548.
XX
PR 16-JAN-1991; 91GB-0000975.
PR 16-JAN-1991; 91GB-0000962.
PR 16-JAN-1991; 91GB-0000963.
PR 16-JAN-1991; 91GB-0000974.
XX
(NICA-) JAPANESE FOUND CANCER RES.
PA (UJO) UNIV JOHNS HOPKINS.
PA (UTAH) UNIV UTAH.
PA (ZENE) ZENECA LTD.
XX
PI Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ;
PI Joslyn G, Kinzler K, Markham A, Nakamura Y, Thliveris A;
PI Vogelstein B, White RL;
XX
DR WPI; 1997-372053/34.
DR N-PSDB; AAT95538.
XX
Cancer diagnosis - by detecting mutation(s) in adenomatous polyposis
PT coli gene
XX
```





Db 2553 khssalp--rvstwrtsagssilssassekseksedkhvnsisgkqskengvsakg 2610  
 QY 1957 SSSSLSSHYE--TPSDALEVISPASSAPPQEKLOTQOPEVVVKANQANDTRQYEGPLH 2015  
 Db 2611 twrkikeneifpntstqvsagatngaesktliymapavsk---tedvvrriedcpin 2667  
 QY 2016 HYRPOQESPSQOQLPP-----SSQAEGMGVPRTHRLITLADHI 2055  
 Db 2668 npr-----srsptgtppvidsvsekanpnikdskdnqakqnvngsvpm--rtvglenri 2722  
 QY 2056 COIITQDFARNQVSSQTPOQPT--STFQNGPSALVST---PVRTKTSNRYSPESQAQSV 2110  
 Db 2723 tsfigvd-----apdqgkteikpgnupvpvsetnespivert-----pfssssss 2768  
 QY 2111 HHQRP-----GSRVSPENLVDSRGRSPKSPERSHVSSEPEPISPPQVPVH--EKQDS 2164  
 Db 2769 khsspsgtvaarvtfnfny-----npsprksasadsarpsqip-tpvnnmtkkrrds 2818  
 RESULT 15  
 AAW38370  
 ID AAW38370 standard; Protein; 2843 AA.  
 AC AAW38370;  
 XX  
 DT 08-APR-1998 (first entry)  
 DE Human adenomatous Polyposis coli gene product.  
 XX  
 KW Human; adenomatous Polyposis coli; APC; diagnosis; prognosis;  
 KW neoplastic tissue; tumour tissue; tumour repressor; mutation;  
 KW sporadic colorectal cancer; detection.  
 XX  
 OS Homo sapiens.  
 XX  
 FN US5691454-A.  
 XX  
 PD 25-NOV-1997.  
 XX  
 PF 25-MAY-1995; 95US-0452654.  
 XX  
 PR 16-JAN-1991; 91GB-0000975.  
 PR 16-JAN-1991; 91GB-0000962.  
 PR 16-JAN-1991; 91GB-0000963.  
 PR 16-JAN-1991; 91GB-0000974.  
 XX  
 PA (CANC-) CANCER INST.  
 PA (ICIL) IMPERIAL CHEM IND PLC.  
 PA (UYJO) UNIV JOHNS HOPKINS.  
 PA (UTAH) UNIV UTAH.  
 XX  
 PI Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ;  
 PI Joslyn G, Kinzler K, Markham AF, Nakamura Y, Thliveris A;  
 PI Vogelstein B, White RL;  
 XX  
 DR WPI; 1998-017712/02.  
 DR N-PSDB; AAT96153.  
 XX  
 PT Antibodies to normal and mutant adenomatous polyposis coli proteins  
 PT - useful for detecting genetic predisposition to cancer  
 XX  
 PS Example 1; Columns 33-52; 107pp; English.  
 CC  
 CC The present sequence is the human adenomatous Polyposis coli (APC)  
 CC gene product, which was used in the development of a novel method  
 CC of diagnosing or prognosing an APC gene associated neoplastic  
 CC tissue. The method comprises comparing APC gene coding sequences or  
 CC mRNA in a tumour tissue, to APC gene coding sequences or mRNA in a  
 CC non-neoplastic tissue, where a difference indicates an APC gene  
 CC associated neoplasia of the tumour tissue.  
 CC APC is a tumour repressor expressed in most normal tissues. APC  
 CC mutations are found in familial adenomatous polyposis and sporadic

CC colorectal cancer patients. The method enables mutations to be  
 CC detected to provide an indication of predisposition to cancer.  
 XX  
 SQ Sequence 2843 AA;  
 Query Match 2.9%; Score 368; DB 19; Length 2843;  
 Best Local Similarity 19.4%; Pred. No. 1.8e-11;  
 Matches 418; Conservative 281; Mismatches 797; Indels 664; Gaps 98;  
 QY 347 PEIRKQEQE--RFQVQORGAGLSATIARSEH-----EISEIIDGLSQENNEK 395  
 Db 981 psiesyddesdkcsygyypadiakhisaubmdndgeldtldpynyslkysdeqlnsgr 1040  
 QY 396 QMRQLSVIPPMFQAEQRRVKFINWGLMEDPMKYKDRQFMNVWTDHE--KEIFKDKFI 453  
 Db 1041 q-----spqnerwarpkhiledikseqgrsrdsttppvytestddkhl 1087  
 QY 454 QHPKNFGLIASYLERKSPVDCVLYYLTKKNENYKALVRRNYKRRGRCQOIARPSOEK 513  
 Db 1088 kfqbhfg-----qqecvspy-----rsrgangsetnrvsgnhginqnvsqleq- 1132  
 QY 514 VEKEEDKAETK--EKEKKDEEKEDESKENTKEKDKIDGTAEETEERQOAPRG 571  
 Db 1133 -ddyeddkptnyseryseeeqheeeerptnysikyn-eehrhvdqpidyslkayatdipss 1190  
 QY 572 R-----KTANSQRRKGRITRSMTEAAAAA-----AAAAATEEPPLPPP 614  
 Db 1191 qkqfsfsssgsgsktehmssssentstpsnaktqnglhpssadsrgsqgkaatck 1250  
 QY 615 PEPISTPEVET-----SRWT-----EEMEVAKKGLVEHGRNMAAIKAM--- 653  
 Db 1251 vssingetiqticyvedtpicfscrsissisaedeigcngttqeadsantliaekgk 1310  
 QY 654 VGTKSE-----AQCKNFYFYKRRHNLNLLQHK-----QKTSR 688  
 Db 1311 igtrsaedpvsavpavaghprrtkssriggssalsesarhkaavefpagksksgagtpk 1370  
 QY 689 KPREE-----RDVSOQESVASTVSAQEDIEDIASNEEENPEDESEAVEAKPSE--DSPEN 740  
 Db 1371 sppehyvqetplmfsrtctsv--ssidsfestsiasvsgepcsmvgsilspdsldpdpqg 1429  
 QY 741 A--TSRGNTPEAVELEPTTAPSTPSLAVPTKPADESVETQVNDISAEATAEQMDV 798  
 Db 1430 tmppsrsktppp---ppqtaqtkevknkaptaeakresgpkqaavnaavq-rvqvlpda 1485  
 QY 799 DQEQHSAEE---CSVC-----DPPPATKADSDVDVEVR---VPEN-HASKVEGDN 840  
 Db 1486 dtllhfatestpdgfsccsslsalsldpfiqk---dveirlmpvpqengdngneteseq 1541  
 QY 841 TKE--RDLDRASEKVEPRDELDVVAQINAQRPBPQSDNDSSATCSADEVD----- 890  
 Db 1542 pkesnenqeakektidsekdl-----ddsd-----dddieileeciis 1581  
 QY 891 -----GEPERQRMF-----PMDSKPSLNPRTGSLVSSPLKPNPLDLPLQAHRAVI 937  
 Db 1582 amptkssrkqkpaqtasklpppvarkpsql-pvykllpsq---nrl---qpqkhvst 1633  
 QY 938 P-----PMVSC---TPCNIPITGTPVSGYAL-----YQRHKAMHE 969  
 Db 1634 pgddmrvyvcgtplnfstatsltditiesppnelaagevrgvgagsegekrtdtipte 1693  
 QY 970 SAL-----LFEQRQREQIDLECRSTSPCGTSKSPNREWEVL----- 1007  
 Db 1694 grstdeaggtktsvtipeiddnkaeeegdliaecinsampkghkpfvrvkkindqvqqa 1753  
 QY 1008 ---OPAPHQLITNLPQGVRLPTTRTPRPPLI-----p 1038  
 Db 1754 sassaapnk---nqldgkklkktspvkipqnteytrvrknadsknnlnaervfsdnk 1810  
 QY 1039 SSKTTVASEKPSF-----IMGSGIS-----QGTPTGTYLTHNQASYTOETP 1079  
 Db 1811 skkqlknnskdndklnpnnedrvrgsfafspshytpiegtp--ycfsrnd----- 1860

Qy 1080 KPSVGSISLGLPROQESAKSATLPYIKOEESPRSONSQOEGLLVRAQHEGVWRTAGAI 1139  
Db 1861 --slsldfdddvdlrkaelrkakenkesakvtshte--ltsnqgsa--nktqai 1913  
Qy 1140 QEGSITRGTTPSKISVESIPSLRSITQGTALPOTGIPTALVK-----G 1185  
Db 1914 akqpinrggpkliqkgs-----tfqsskdipdrgaatdekliqfaientpvcfshns 1967  
Qy 1186 SISRM-----PIEDSPKREAAKSGHVIYEGKSGHI----- 1219  
Db 1968 slsldidqennkenepiketepdsqgepskpaasyapksfhvedtpvcfsrns 2027  
Qy 1220 --LSYDN-----IKNAREGTRSPRTAHEISLKRYSYESVEGNIKOGMSRSPVSAPL 1269  
Db 2028 ssldisdellqecissampkkkpsrlkgdnekhspnmngilgedltldkdiqrd 2087  
Qy 1270 EGLICRALPRGSPHDLKERVLSGSIQGTTPRATTSFEDGLKYPQIKRESPIRAFE 1329  
Db 2088 se---hgispdsenfkaIQegansivsslhqaaaac-----lsrqassdsllslk 2139  
Qy 1330 CAITKGRPYDGIITIKEMGRSIHPIRQDILTQESRKTPEVVQSTRPIEISISQGTPIK 1389  
Db 2140 sgislgpf-----hltpdeekpftcnkgprilk-----pgekstletkk 2180  
Qy 1390 FDNNS-GOSAIKHNKSLITGP-----SKLSRGM-PPLEIVPENIKVVERKGYEDVKAGET 1443  
Db 2181 lseskgikggkkykslitgkvrnsesigqmkqplq---anmpsirg----- 2227  
Qy 1444 VRSRHTSVSSGSPVSLSTLHEAPKAQLSPGIYDDTSARRTPVSQNTMSRGSPPMNRTS 1503  
Db 2228 -----rtmih-----ipgv-----rnsstspvskkkgp 2251  
Qy 1504 DVTIPPNNKSTNHERKSLTTP-QRESIPAK-SPVPDGVDPVVSHPSPDPHHRGSTAGEVYW 1561  
Db 2252 plktpaakspegatatsprgakpskkselsvarqtqigsskapsrgsr----- 2305  
Qy 1562 SHLPTQLDPAMPFRHALDPAAAYLFORQLSPTEGYPYSQVQLYAMENTROTILNDYITSQ 1621  
Db 2306 -----dstpsrpaqpl----- 2319  
Qy 1622 QMOYNLRPDVA---RGLSPREOPLGLPYPATRGIDITNMPTTLVPHPGGTSTPPMDRI 1678  
Db 2320 plqspgrnsispgngslspnk-----lsqprt-sspsstasksgsgkm 2364  
Qy 1679 TIPTQTITPPRPNYNSASPCGHTHLAAAASAEERERERE-----KERER 1726  
Db 2365 sy-----tspgrmqsqnit--kqtglknassiprsesaskglngmngngankkvel 2416  
Qy 1727 ERIAAASDLYLRPGSPGCRPGSHGVRSPPSVRTQETMLQORPSVFOGTNGTSTVTP 1786  
Db 2417 srmsstkss-----gsesd-----rserpvlvrqstfikeaps----- 2449  
Qy 1787 LDPTAQLRIMPLPAGGPSISQGLPASRYNTAADLALVDAASAPQMDVSKTESKHEA 1846  
Db 2450 --ptlr-rkleesafesls---psrpsaptrsqadtplslpaldmsls-thssvqag 2502  
Qy 1847 A--RLEENLRSRAAASVSEQOLEKTELEKRSVQCLYTSAPPSGKPPQHSSVYSEAG 1904  
Db 2503 gwrklpnl---sptieyndgrpakrhdia-----shsesprlpnrsgtwkrehs 2552  
Qy 1905 KDGPPPKPSRYEELRTRGKTTITPAAN-----FIDVITRITQASDKDARERGSSSD 1956  
Db 2553 khssslp--rvstwrtrgssslsassesekaksedekhvnsisgtkqskengvsaakg 2610  
Qy 1957 SSSLSLHRYE-TPSDAIEVISPASSAPQEKLTQVPEVVKANQENPTTROYEGPLH 2015  
Db 2611 twrkikeneftsptnststqvtssgatngaesktliymapavsk---tedvwriedcpin 2667  
Qy 2016 HYRPOQESPSQOQLPP-----SSQAEQMGQVPRTHRLITLADHI 2055  
Db 2668 npr-----sgrsptgntppvidsvsekanpnikdkdnqakqngvngsvpm---rtvglenrl 2722

Qy 2056 CQIITQDFARNQVSSQTPQOPPT--STFONSPSALVST---PVRTKTSNRYSPESQAOVS 2110  
Db 2723 tsfiqvd-----apdqgtkqpnnpvpysetnespivert-----pfssssss 2768  
Qy 2111 HHQRP-----GSRVSPENLVDSRGRSPKSPERSHVSSEPIEPISPPQVPPVH--EKQDS 2164  
Db 2769 khspsgtvaarvtpfny-----npsprksadatsarpsqip-tpvnnntkkrds 2818  
RESULT 16  
AAR63507  
ID AAR63507 standard; Protein; 2860 AA.  
XX  
XX AAR63507;  
DT 30-MAY-1995 (first entry)  
XX  
DE Adenomatous polyposis coli tumour repressor.  
XX  
XX Adenomatous polyposis coli; tumour repressor; Gardner's syndrome;  
KW familial adenomatous polyposis; cancer diagnosis and prognosis;  
KW tumorigenesis suppression.  
XX  
OS Homo sapiens.  
XX  
XX US5352775-A.  
PN  
PD 04-OCT-1994.  
XX  
PF 08-AUG-1991; 91US-0741940.  
XX  
XX 16-JAN-1991; 91GB-0000962.  
PR 16-JAN-1991; 91GB-0000963.  
PR 16-JAN-1991; 91GB-0000974.  
PR 16-JAN-1991; 91GB-0000975.  
XX  
PA (CANC-) CANCER INST.  
PA (ICIL ) IMPERIAL CHEM IND.  
PA (UJO ) UNIV. JOHNS HOPKINS.  
PA (UTAH ) UNIV UTAH.  
XX  
PI Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ;  
PI Joslyn G, Kinzler K, Markham AF, Nakamura Y, Thliveris A;  
PI Vogelstein B, White RL;  
XX  
DR WPI; 1994-316233/39.  
DR N-PSDB; AAR63507.  
XX  
PT New human adenomatous polyposis coli DNA encoding tumour  
PT repressor - and derived primers and probes for diagnosis,  
PT prognosis and treatment of cancer  
XX  
PS Claim 3; Fig 7; 113pp; English.  
XX  
CC AAR63507 is a cDNA isolated from the human adenomatous polyposis  
CC coli (APC) gene, it encodes the tumour repressors described in AAR63507  
CC and AAR63508. Determination of alterations in APC or its expression  
CC products, can be used for the diagnosis and prognosis of cancer e.g.  
CC colorectal, lung and breast tumours; and for determining predisposition  
CC to certain cancers such as familial adenomatous polyposis (FAP) and  
CC Gardner's syndrome. The wild type APC gene (or a part of it) can be  
CC used therapeutically to restore gene function, while primers and probes  
CC derived from the cDNA (AAR63507-400 and AAR63507-568) can be used to  
CC detect mutations. Also APC proteins or analogues can be administered to  
CC compensate for a defective gene, and epithelial cells, or transgenic  
CC animals carrying a mutated APC allele are useful for detecting  
CC therapeutic agents able to suppress tumorigenesis.  
SQ Sequence 2860 AA;  
Query Match 2.9%; Score 368; DB 15; Length 2860;  
Best Local Similarity 19.4%; Pred. No. 1.9e-11;



## RESULT 17

AAR58634  
ID AAR58634 standard; Protein; 2843 AA.

AC AAR58634;

DT 21-JUN-1995 (first entry)

DE Adenomatous polyposis coli protein (APC).

XX Adenomatous polyposis coli protein gene;

KW familial adenomatous polyposis; colorectal tumor; adenoma.

XX Homo sapiens.

XX W09421814-A.

XX 29-SEP-1994.

XX 21-MAR-1994; 94WO-US02987.

XX 19-MAR-1993; 93US-0034850.

XX (UUYO) UNIV JOHNS HOPKINS.

XX Hill DE, Johnson KA, Kinzler KW, Vogelstein B;

XX WPI; 1994-317033/39.

DR Q-PSDB: Q70633.

PT Antibodies to adenomatous polyposis coli protein - are used for

PT detecting mutations in the APC gene for predicting

PT pre-disposition to cancer, partic. colon cancer

XX Claim 1; Page 46; 8lpp; English.

XX Antibodies prepared to this protein are used for detecting mutations  
CC in the APC gene for predicting pre-disposition to cancer, particularly  
CC colon cancer. The Abs can detect germ line or somatic mutations  
CC indicating a predisposition to colon cancer and possibly gastric,  
CC oesophageal, pancreatic or small cell lung cancers.

XX Sequence 2843 AA;

Query Match 2.98; Score 366; DB 15; Length 2843;

Best Local Similarity 19.48; Pred. No. 2.4e-11;

Matches 418; Conservative 280; Mismatches 798; Indels 664; Gaps 98;

Qy 347 PEIRKQROOE-RFQRVGQAGLSATARSEH-----EISEIDGLSQENNEK 395

Db 981 psiesyddeskcscvgyqpadlahkhsanhdndgelldtpinsikysdeqlnsgr 1040

Qy 396 QMRQLSVIPMMFDAQRRVRFKFINNGLMEDPMKYKDQPMNYWTQHE--KEIFKDKFI 453

Db 1041 q-----spsqnerwarpkhiledikgseqrgsrngstpyvtestddkhl 1087

Qy 454 QHPKNFGLIASYLERKSPVDCVLYLYLTKKKENYKALVRNRYKRRGNQOIARPSQEEK 513

Db 1088 kfqphfg-----qqecvspy-----rsrgangsetnrvngnhlnqnvqsldcqe- 1132

Qy 514 VEEKEEDKAETK--EEEKDDEEKDEKEDSKENTKEKDKIDGTAEETEEREQATPRG 571

Db 1133 -dyeddktptnysryseeeehheerptnysikyn-eekrhvdpdiyslkysktdipss 1190

Qy 572 R-----KTANSQGRKGRTRTSMWNEAAAS-----AAAAATEEPPPLPPP 614

Db 1191 qkqsfstkskssggsktelmssttsntpsnkrqnlhpsaaqsrsgpckaack 1250

Qy 615 PEPISTEPVET-----SRWT-----EEEMEYAKKGLVEHGRNWAATAKM--- 653

Db 1251 vssinqetigtvcyvedtpicfscssissaaedgicnqttqeadsantlqlaeikgk 1310

Qy 654 VGTKSE-----AOCKNFYFNKRRHLDNLLOQHK-----OQTSTR 688

Db 1311 igtrsaedpvsevpavsqhprtkksrlqgsslsesahrkavfsgakspksqagtpk 1370

Qy 689 KPREE-----RDVSCQESVASTVSAQEDIEDIASNEENPEDSEVEAVKPSE--DSPEN 740

Db 1371 spbehvqetplmfsrctsv-ssidsfesrslaasvqpcsgmlsgilspdlpdsppq 1429

Qy 741 A--TSRGNTPEAVELEPTTETAPSTPSLAVPSTKPADESVETQVNDISIAETAEQMDV 798

Db 1430 tmppsrsktppp---ppqtaqtrelpkntaekresgpkgaavnaqv-rvqvlpda 1485

Qy 799 DQOQHSABE---GSVC-----DPPATKADSVDEVR---VPEN-HASKVEDN 840

Db 1486 dtllhfatestpdgfcscsslsalsldpfigk---dvelrmpvqendngneteseq 1541

Qy 841 TKE--RDLDRASEKVEPRDELVVAQOINAOPEPQSDNSDSSATCSADEVD----- 890

Db 1542 pkesnenqeakektidsekdl-----ddsd-----dddieileecilis 1581

Qy 891 -----GEPERQRMF-----PMDSKPSLLNPTGSIILVSSPLKPNPLDLQLQHRAVI 937

Db 1582 amptksrkgkpaqtasklpppvarkpsqi-pvykilpsq---nrl---qpqkhvst 1633

Qy 938 P---PMVSC---TPCNIPICTPVSGYAL-----YORHIKAMHE 969

Db 1634 pgddmrvyvcvgtpinfstatsldtiesppnelaagevrggaqggefekrdtipte 1693

Qy 970 SAL-----LEEQRQREQIDLECRSSTSPCGTSKSPNREWEVL----- 1007

Db 1694 grstdeagggktsvtipelddnkaeegdilaecinsampkgkshkpfvrkkmindqvqga 1753

Qy 1008 ---QPAPHQLITNLPEGVRLFTTRTPPPPLI-----P 1038

Db 1754 sassaapnk---nqldgkklkxtpspkvpipqnteytrvrknadsknlnaervfsdndk 1810

Qy 1039 SSKTTVAASEKPSF-----IMGGSTF-----OQTGTYLTSHNOASTQETP 1079

Db 1811 skknlknmskdfndklnnedrvrgsfafdsphhytpiegtp-ycfsrnd----- 1860

Qy 1080 KPSVGSISLGLPROQESAKSATLPYIKOEFSRPSQNSQPEGLLVRAQHEGVVRGTAGAI 1139

Db 1861 --slsldfddddvdlrsrekaelrkakenkesakvtshte--ltsnqgsa---nktqai 1913

Qy 1140 QEGSITRGTPTSKISVESIPSLRGSITQGTALPOTGTPTEALVK-----G 1185

Db 1914 akqplnrgqpkpiliqkgs-----tfpqskdipdrgaatdekigqnfalentaipvcfshns 1967

Qy 1186 SISRM-----PIEDSSPEKGRGEEAASKGHVYEGKSGHI----- 1219

Db 1968 slsalsdidqennnnkenepiketepdsqgqpskpaqsgyapkfshvedtpvcfsrns 2027

Qy 1220 --LSYDN-----IKNAREGTRSPRTAHEISLKRYSVESVEGNIKQGMRESPEVSAPL 1269

Db 2028 sslsldseddlilqecissampkkkpsrlkgdnekhspnmnggilgedltldldkigrp 2087

Qy 1270 EGLICRALPGCSPHSDLKERTVLSGSIWQGTPTATTESFEDGLKVPKQIKRESPIRAFE 1329

Db 2088 se---hglspdsenfwdkalgqegansivsslhqaaaac-----lsrqassdsdsilsk 2139

Qy 1330 GAITKGPYDGTITTIKEMGRSIEHPQDILTQESRKTPEVVSQSTRPIEISISQGTPIK 1389

Db 2140 sgislgpf-----hltqdeekpfstnkgprllk-----pgekstlekk 2180

Qy 1390 PDNNS-GQSAIKHNKSLITGP-----SKLSKGM-PPLEIVPENIKVVERGYEDVKAGET 1443

Db 2181 ieseskigkggkvykslitgkvrnsseisgmkpqlq---annpmsrg----- 2227

Qy 1444 VRSRHTSVVSSGSPSVLSTLHEAPKAQLSPGIYDDTSARRTPVSVQNTMTRSGSPMMNRTS 1503

Db 2228 -----rtmih-----ipgv-----rnsststspvskkgp 2251





```

OS Homo sapiens.
XX US5352775-A.
XX
XX PD 04-OCT-1994.
XX
XX PF 08-AUG-1991; 91US-0741940.
XX
XX PR 16-JAN-1991; 91GB-0000962.
XX
XX PR 16-JAN-1991; 91GB-0000963.
XX
XX PR 16-JAN-1991; 91GB-0000974.
XX
XX PR 16-JAN-1991; 91GB-0000975.
XX
XX (CANC-) CANCER INST.
XX
XX PA (ICIL) IMPERIAL CHEM IND.
XX
XX PA (UXJO) UNIV. JOHNS HOPKINS.
XX
XX PA (UTAH) UNIV UTAH.
XX
XX
XX PI Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ;
XX PI Joslyn G, Kinzler K, Markham AF, Nakamura Y, Thliveris A;
XX PI Vogelstein B, White RL;
XX
XX WPI; 1994-316233/39.
XX
XX DR N-PSDB; AAQ72297.
XX
XX
XX PT New human adenomatous polyposis coli DNA encoding tumour
XX PT repressor - and derived primers and probes for diagnosis,
XX PT prognosis and treatment of cancer
XX
XX PS Claim 3; Columns 71-84; 113pp; English.
XX
XX CC AAQ72297 is a cDNA isolated from the human adenomatous polyposis
XX CC coli (APC) gene, it encodes the tumour repressors described in AAR63507
XX CC and AAR63508. Determination of alterations in APC or its expression
XX CC products, can be used for the diagnosis and prognosis of cancer e.g.
XX CC colorectal, lung and breast tumours; and for determining predisposition
XX CC to certain cancers such as familial adenomatous polyposis (FAP) and
XX CC Gardner's syndrome. The wild type APC gene (or a part of it) can be
XX CC used therapeutically to restore gene function, while primers and probes
XX CC derived from the cDNA (AAQ72333-400 and AAQ72541-568) can be used to
XX CC detect mutations. Also APC proteins or analogues can be administered to
XX CC compensate for a defective gene, and epithelial cells, or transgenic
XX CC animals carrying a mutated APC allele are useful for detecting
XX CC therapeutic agents able to suppress tumorigenesis.
XX
XX SQ Sequence 2842 AA;

Query Match 2.9%; Score 365; DB 15; Length 2842;
Best Local Similarity 19.0%; Pred. No. 2,7e-11;
Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

QY 347 PEIRKQROQE-RQQRVGQRCAGISATIAHSEH-----BISEIIDLSEQENNEK 395
DB 980 psiesyseddeskscsygqypadiahkhsanhdnddgeltdtpnyslkysdeqlnsgr
QY 396 QMRQLSVIPPMFADQORRVFVFNNGMLMDPMKVYKDRQPMNVYTDHE--KEIFKDKFI 453
DB 1040 q-----spqnerwarphkhiiedeikqseqqrnsqntypvytestdkhl 1086
QY 454 QHPKNFGLIASYLERKSVDPCLVYYITKKNENYKALVRRNYGKRGRNQOIAQPSOEK 513
DB 1087 kfqbhf-----qgcvcvpy-----rsrgangsetnrvsnhginqvsgslqce- 1131
QY 514 VEEKEEDKAETKK--EEKKDEEKEDKSKENTKEKDKIDGTAEETEERQANPRG 571
DB 1132 -doyeddtkptnysryseeqneeeertnysikyn-eehrhvdqpdlyskyatdipss 1189
QY 572 R-----KTANSQRRKGRITRSMTEAAAS-----AAAAATEEPPLPLPP 614
DB 1190 qkqfsfkskssgsksthemsssentstpsnaknqnlhpsaqsqrsgqpkatck 1249
QY 615 PEPISSTEPVET-----SRWTEEMEVAKKGLVEHGRN-----WAAIAKM 653

```



```
Db 2227 -----rtmih-----ipgv-----rnssstspvskgpp 2251
Qy 1505 VTIPNKTSTHERKSTLTPT-QRESIPAK-SPVPGVDVWVSHSPDPHHRGSTAGEVYWS 1562
Db 2252 lktbaskspsegqtattspqakpsvkselapqtsiqgsskapsrsgs----- 2304
Qy 1563 HLPQLDPAMPFHRALDPAAAYLFQRLSPTPGPSQYQIYAMENTROTILNDIYTSQ 1622
Db 2305 -----dstpapaqpl-----srp 2319
Qy 1623 MQVNLRPDVA---RGLSPREQPLGLPYPATRGIIIDLTMNPPTILVPHPGG*STPPMDRIT 1679
Db 2320 iqsggrnsisprngisppnk-----lsqiprt-spsstaskssgsgkms 2364
Qy 1660 YIPGTQITFPPRPYNSASMSGPHPTLAAAASAEERERERE-----KERERE 1727
Db 2365 y-----tspgrmsqqnlt--kqtglsknassipresaskglnqnmngngaukxkvels 2416
Qy 1728 RIAAASDLILRPGSEQPGRGSHGVYRSPSPSVTQETMLQORPSVFGTNGTSVITPL 1787
Db 2417 rmsstkas-----gseqd-----rserpvivrgstfikeaps----- 2448
Qy 1788 DPTAQLRIMPLPAGGPGSISQGLPASRYNTAADALAALVDAASAPQMDVSKTKSKHEAA 1847
Db 2449 -ptlr-kleesafesis---psrpsatrsqaqtpvlspslpmisls-thsvqagg 2502
Qy 1848 --RLEENLRSAAVSEQQLEQKTLVEKRSVQCLTSSAFFSGKPPQHSVVVSEAGK 1905
Db 2503 wrklppnl---sptieyndgrpakrhdia-----shsesparlpinrgstkrehsk 2552
Qy 1906 DKGPPPKSYEEELTRGKTITTAAN-----FIDVIITRQIASDKDARERSQSDDS 1957
Db 2553 hssalp--rvatwrtgsssilasessesekaksedekhvnsisgtkqskengvsaagt 2610
Qy 1958 SSSLSHRYE-TPSDATEIVISPASSPAPPQEKLOTYQPEVVKANQAENDPTRYEGPLHH 2016
Db 2611 wrklkenefstnstsqtvsngaekskliymgavask-----tedvwriedcpinn 2667
Qy 2017 YRQOESPSQOQLPP-----SSQAEQMGQVPRTHRLITLADHIC 2056
Db 2668 pr---sgrsptgnppvidsvsekanpnikskdnqakqngvngsvpm--rtvglenrln 2722
Qy 2057 QIITQDFARNQVSSQTP-QQPTSTFQNSPSALYS-TPVRKTKSNRYSPESQAOVHHQR 2114
Db 2723 sflqvdapdqgtktpqgnnpvpvsetnessivertpfssssskhspsgtva----- 2777
Qy 2115 PGRSVSPENLVKSGRSPGKSPERSHVSSEPYEPISPPQVPVVH--EKQDS 2164
Db 2778 --arvtfny-----npsprkssadtsarpsqip-tpvnnntkkdds 2817

RESULT 20
AAW76140
ID AAW76140 standard; Protein; 2843 AA.
XX
AC AAW76140;
XX
XX 23-NOV-1998 (first entry)
XX
XX Human APC protein #1.
XX
XX Familial adenomatous polyposis coli; APC; tumour suppressor; therapy;
KW chromosome 5q21; tumorigenesis; retinoblastoma; colorectal tumour;
KW FAP; Gardner's Syndrome; GS; predisposition.
XX
XX Homo sapiens.
XX
XX US578366-A.
XX
XX 21-JUL-1998.
XX
XX 25-MAY-1995; 9505-0452655.
XX
```

```
PR 16-JAN-1991; 91GB-0000975.
PR 16-JAN-1991; 91GB-0000962.
PR 16-JAN-1991; 91GB-0000963.
PR 16-JAN-1991; 91GB-0000974.
XX
PA (CANC-) CANCER INST.
PA (UYJO) UNIV JOHNS HOPKINS.
PA (UTAH) UNIV UTAH.
PA (ZENE) ZENEGA PHARM.
XX
PI Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ;
PI Joslyn G, Kinzler K, Markham AF, Nakamura Y, Thliveris A;
PI Vogelstein B, White RL;
XX
DR WPI; 1998-427100/36.
DR N-PSDB; AAV56447.
XX
PT Adenomatous polyposis coli protein - useful in the treatment of
PT cancers associated with mutation(s) on human chromosome 5q21
XX
PS Disclosure; Column 41-54; 102pp; English.
XX
CC This sequence represents a human familial adenomatous polyposis coli
CC (APC) protein from clone DP2.5. The gene for the protein is present on
CC human chromosome 5q21 and is also referred to as adenomatous polyposis
CC coli gene. It is a tumour suppressor gene, and mutations in this gene
CC have been associated with tumorigenesis in retinoblastoma and colorectal
CC tumours, and especially familial adenomatous polyposis (FAP) and
CC Gardner's Syndrome (GS). The protein can be used in therapy to replace
CC lack of native functional protein and the nucleic acids can be used for
CC gene therapy. The nucleic acids that encode them can also be used as
CC probes and primers in detection of the cancers and predisposition to it.
XX
SQ Sequence 2843 AA;
```

```
Query Match 2.9%; Score 365; DB 19; Length 2843;
Best Local Similarity 19.0%; Pred. No. 2.7e-11;
Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

Qy 347 PEIRKQREQE-RQRVQRCAGLSATARSEH-----EISEIIDGLSEQENNEK 395
Db 981 pslesyseddesktcsygyqpadlahkhsanhdmdndgeltdtpinsyksydesqlnsgr 1040
Qy 396 QMRQLSVTPPMFADAEQRRVFINNGLMEPMKYKDRQFMNVTDHE--KEIFKDFI 453
Db 1041 q-----spsqnerwarphkhiiedeikseqqrngsttypvtestddkhl 1087
Qy 454 QHPKNFGLIASYLERKSVDPCLVLYYYITKKNENYKALVRRNYKRRNQIARPSQEEK 513
Db 1088 kfphfg-----qgecvsp-----rsrgangsetnrvsnhginqnvsgslcqe- 1132
Qy 514 VEEKEEDKAEKTEKK--EEKKDEEKDEKEDSKENTKEKIDGTAEETEREQATRG 571
Db 1133 -ddyeddkptnyseryseehheerptnysikyn-eehrhvdqpidysikyatdips 1190
Qy 572 R-----KTANSQGRKGRITRSMTEAAAS-----AAAAATEEPPLPPP 614
Db 1191 qkqsfsksssgssktehmssssentstpsnakrnqlhpsaqrsqgqkaack 1250
Qy 615 PEPISTEPVET-----SRWTEEMEVAKKGLVEHGRN-----WAAIAKM 653
Db 1251 vssinqetiqticyvedpicfscsslsllsaedeigcncqtqeadsantliaelke 1310
Qy 654 VGTKSE-----AQCKNFYFNKRRHLDNLLQOHK-----QKTSR 688
Db 1311 igtrsaedpvsevpavsqhprtksrliqgsslsasathkavefssgskspksagatpk 1370
Qy 689 KPREE-----RDVSOCSVASTVSAQDEDEIDIASNEENEPDESVEAVKPSE--DSPEN 740
Db 1371 sppenyvqetplmfsrcstsv--ssldsfesrslasvqpcsgmvsnglispdlpdpqg 1429
Qy 741 A--TSRGNTPEAVELEPTTETAPSTSPSLAVPSTKPADESVEVQVNDISAEQMDV 798
```





Db 2450 -ptlrr-rkleeasfesis---psrpsasrtsgaqtvpvlpslpsldmsls--thssvqag 2503  
 QY 1848 --RLEENLRSAASEVSEQQLEQKLEVEKRSVQCLYTSSAFPSKQPPHSSVYSEAGK 1905  
 Db 2504 wrklppnl---sptleyndgrpakrhdiar-----shsesprlpinrgtwwkrehsk 2553  
 QY 1906 DKGPPPKSRYEELRTRKTTITAAAN-----FIDVITRQIASOKDAREGSSQSDS 1957  
 Db 2554 hssslp--rvstwtgsssilssassekaksekdekhvnsigtkqkengqvskgt 2611  
 QY 1958 SSSLSHRYE--TPSDAIEVISPASPAPQEQKLTQYQEVVKANOENPPRQYEGPLHH 2016  
 Db 2612 wrkikenesfptnsgtvsygatngaesktliygmapavsk-----tedvwrwiedcpinn 2668  
 QY 2017 YRQOESPSQQQLPP-----SSQAGSMGQVPRTHRLITLADHIC 2056  
 Db 2669 pr----sgrsptgntppvldsvsekanpnikskdnqakngvngsvpm--rtvlglenrin 2723  
 QY 2057 QIITQDFARNQVSSQTP-QQPPTSTFQNSALYS-TPVRTKTKNRYSPESQAOVHHQR 2114  
 Db 2724 sfliqvdapdkgtelkpgqnnpvpvsetnessivertpfssssskhsspsgtva----- 2778  
 QY 2115 PGRSVSPENLVKSRGSPGSPERSHVSSEPEISPPQVPVYH--EKQDS 2164  
 Db 2779 --arvtpfny-----npsprkssadstsaarpsqip-tpvnnntkkrds 2818

RESULT 22  
 AAB23011  
 ID AAB23011 standard; Proteins; 2843 AA.  
 XX AC AAB23011;  
 XX AC  
 XX DT 16-JAN-2001 (first entry)  
 XX DE Human APC protein (splice variant 1).  
 XX KW APC gene; Adenomatous Polyposis Coli gene; human; chromosome 5q21;  
 KW familial adenomatous polyposis; FAP locus; Gardner's syndrome; GS;  
 KW sporadic tumour; adenoma; carcinoma; cancer; lung; breast; colon; rectum;  
 KW bladder; liver; sarcoma; stomach; prostate; leukaemia; lymphoma;  
 KW tumour suppressor; anti-APC antibody; detection; diagnosis; prognosis;  
 KW genetic predisposition; drug screening; DP2.5; splice variant.  
 XX OS Homo sapiens.  
 XX PN US6114124-A.  
 XX PD 05-SEP-2000.  
 XX PF 25-MAY-1995; 95US-0450582.  
 XX PR 16-JAN-1991; 91GB-0000962.  
 XX PR 16-JAN-1991; 91GB-0000963.  
 XX PR 16-JAN-1991; 91GB-0000974.  
 XX PR 16-JAN-1991; 91GB-0000975.  
 XX PR 08-AUG-1991; 91US-0741940.  
 XX PR 12-AUG-1994; 94US-0289548.  
 XX (ICIL ) IMPERIAL CHEM IND PLC.  
 XX (UYJO ) UNIV JOHNS HOPKINS.  
 XX (UTAH ) UNIV UTAH.  
 XX (CANC-) CANCER INST.  
 XX Carlson M, Groden J, Joslyn G, Kinzler K, Markham AF, Anand R;  
 FI Albertsen H, White RL, Thillveris A, Nakamura Y, Vogelstein B;  
 FI Hedge PJ;  
 XX WPI: 2000-565003/52.  
 XX N-PSDB: AAA93449.  
 XX Detecting Adenomatous Polyposis Coli (APC) protein in a sample for  
 PT diagnosing cancers, involves contacting the sample with antibodies that

PT specifically bind to APC protein and detecting the complex formed -  
 XX Claim 1; Fig 3A-C; 125pp; English.  
 XX The invention relates to a novel method for detecting Adenomatous  
 CC Polyposis Coli (APC) protein in a sample. The method involves  
 CC contacting the sample with antibodies which specifically binds to the  
 CC 2843 amino acid form of the human APC protein, or to a mutant APC  
 CC protein, and detecting an APC-antibody complex. Mutations in the APC  
 CC gene play a role in tumorigenesis, indicating that it is a tumour  
 CC suppressor gene. It is located on chromosome 5q21, which corresponds to  
 CC the FAP (familial adenomatous polyposis) locus. FAP is an autosomal  
 CC dominant inherited disease in which affected individuals develop  
 CC hundreds to thousands of adenomatous polyps in the colon and rectum,  
 CC some of which progress to malignancy. The FAP locus is often found to  
 CC be deleted in sporadic (i.e., non-familial) adenomas and carcinomas, and  
 CC chromosome 5q deletions have also been observed in tumours of the lung,  
 CC breast, colon, rectum, bladder, liver, sarcomas, stomach, and prostate,  
 CC and in leukaemias and lymphomas. Although the FAP locus contains  
 CC several other genes such as FER, TBL1, TBL2, and MCC, it is thought that  
 CC mutations in the APC gene play a key role in the development of FAP and  
 CC sporadic tumours. The method is useful for detecting APC protein and its  
 CC mutant forms in foetal tissue, placental tissue, amniotic fluid, blood,  
 CC serum or a tumour sample. The method is useful for diagnosing or  
 CC prognosing neoplastic tissue, for detecting a genetic predisposition to  
 CC cancer, for detecting germline and somatic alteration of wild-type APC  
 CC genes, and for testing therapeutic agents for the ability to suppress  
 CC tumours. The present sequence represents a 2843 amino acid splice  
 CC variant of the human APC protein. This variant is more abundant than the  
 CC 2742 amino acid variant (AAB23012).  
 XX SQ Sequence 2843 AA;

Query Match 2.98; Score 365; DB 21; Length 2843;  
 Best Local Similarity 19.08; Pred. No. 2.7e-117;  
 Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;  
 QY 347 PEIRKREQQB-RPQRVQGRGAGLSATIRSEH-----EISEIDGLSQENNEK 395  
 Db 981 psieseddeskfcsgyqypadlahkhsanhmddngeldtptnyslkysdeqlnsgr 1040  
 QY 396 QMRSLSVIPPMFADQRRVKFINNGLMEDPMKYKQRFPMVYTWTHE--KEIFKKKFI 453  
 Db 1041 q-----spqnerwarphkileideikqsegrndqsttppvytestddkhl 1087  
 QY 454 QHPKNFGLIASYLERKSVPCVLYLYLTKNENYKALVRRNYGRRNRNQOIARPSQEEK 513  
 Db 1088 kfqphfg-----qkecvsp-----rsrgangsetnrvgshnglnqvsqslqce- 1132  
 QY 514 VEEREEDKAEKTEKK--EKKKDEEKEDEKEDSKENTKEKDKIDGTAEETEEREQATPRG 571  
 Db 1133 -dyeddktptnyseryseeeheeeertnysikyn-eehrhvdqpidyslkvatdipss 1190  
 QY 572 R-----KTANSQGRKGRITRSMTEAAAS-----AAAAATEPEPPPLPPP 614  
 Db 1191 qkqsfstskssggsgsktehmssssststpsnkrqnlhpsaagrsqgqkaatck 1250  
 QY 615 PEPITSTPVEV-----SRWTEEMEYAKKGLVEHGRN-----WAAATKM 653  
 Db 1251 vssinqetigtvcvedtptcfscrossississadeicgcnqtqeadsantlqlaelkek 1310  
 QY 654 VGTKSE-----AQCKNFYFNKRRHNLNLOQHK-----QKTSR 688  
 Db 1311 igtrsaeidpvsavpavsqhqrptksrriqgsslsesarhkavfssgakspsksagtpk 1370  
 QY 689 KPREE-----ROYSQCESVASTVSAQDEDEIASNEENPEDSEVAVKPSK---DSPEN 740  
 Db 1371 sppehyvqetplmfsrctsv-ssldsfesrssiassvqpcsgmvgiispsdlpdpqpg 1429  
 QY 741 A--TSRGNTPAVELEPTTETAPTSTPSLAVPTKPAEDSVETQVNDISAEATQMDV 798  
 Db 1430 tmppsrsktppp---ppqtaqtkevvpknkptaekresgpkgaavnaavq-rvqvlpda 1485





Db 2504 wrklppnl---sptieyndgrprkhrdiar-----shsespsrlpnrsgtwkrehs 2553  
 QY 1906 DKGPPPKRYEELTRGKTTITAN-----FIDVLIITRIQIASDKDAREGQSDDS 1957  
 Db 2554 hsslp--rvstwrtsqssilssassekakaseedekhnvsisgtkqskengvsakgt 2611  
 QY 1958 SSSLSHRYE-TPSDAIEVISPASPAPQEKLOTQYQPEVVKANOAEENDPTRQYEGPLHH 2016  
 Db 2612 wrkikenefspntstqvtvgatngaesktliqmpavsk---tedvwriedcpinn 2668  
 QY 2017 YRQOESPSQOQLPP-----SSQAGMGQVPTTHRLITLIADHIC 2056  
 Db 2669 pr---sgrstgntppvidsvsekanpnkdsdnqakngvngsvpm---rtvglienrln 2723  
 QY 2057 QIITQDFARNQVSSQTP-QQPTSTFQNSPALVS-TPVRTKTSNRYSPESQAOQSVHQR 2114  
 Db 2724 sfiqvdapdqgkeltkpgnnpvsetnessivertfssssskhspsgtva----- 2778  
 QY 2115 PGRSVSPENLVDKSGRPGKSPERSHVSSEPYEPISPPQVPPVH--BKQDS 2164  
 Db 2779 --arvtpfny-----npsprkssadstsrpsqip-tpvnnntkkrds 2818

RESULT 24  
 AAY70304  
 ID AAY70304 standard; Protein; 2973 AA.  
 XX AC AAY70304;  
 XX DE  
 DT 06-JUN-2000 (first entry)  
 XX DE  
 KW Human Tcf-4E protein; transcription factor; beta-catenin.  
 KW Adenomatous polyposis coli; APC; transcriptional activation;  
 KW Tcf responsive reporter gene; APC transcription regulatory pathway;  
 KW familial adenomatous polyposis; FAP; cancer; colorectal; thyroid; brain;  
 KW medulloblastoma; breast; head; neck; desmoid tumour; osteoma;  
 KW cytostatic.  
 XX OS Homo sapiens.  
 XX PN WO200011195-A1.  
 XX PD  
 XX PF 02-MAR-2000.  
 XX PR 20-AUG-1999; 99WO-US18774.  
 XX PR 20-AUG-1998; 98US-0136605.  
 XX PA (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 XX PI He T, Vogelstein B, Kinzler KW;  
 XX WPI: 2000-237657/20.  
 XX PT Determining wild-type adenomatous polyposis coli protein for diagnosing  
 XX PT cancer comprises introducing a Tcf responsive reporter gene having  
 XX PT upstream sequences of c-MYC into a cell  
 XX PS Disclosure; Page 58-65; 70pp; English.  
 XX CC The present sequence is a protein involved in cancer diagnosis  
 XX CC associated with APC or beta-catenin mutations. Human Tcf-4E protein  
 XX CC expressed in colorectal epithelium transactivates transcription  
 XX CC when associated with beta-catenin. Adenomatous polyposis coli (APC)  
 XX CC regulates this transcriptional activation, at least in part by binding  
 XX CC to beta-catenin. Determining wild-type APC protein for diagnosing cancer  
 XX CC comprises introducing a Tcf responsive reporter gene having upstream  
 XX CC sequences of c-MYC into a cell and measuring transcription of the  
 XX CC reporter gene. The candidate drug identified is useful for treating  
 XX CC familial adenomatous polyposis patients with APC or beta-catenin  
 XX CC mutations and patients with increased risk of developing cancers such as

CC colorectal, thyroid, brain, medulloblastoma, desmoid tumour, osteoma,  
 CC breast, head and neck.  
 XX SQ Sequence 2973 AA;  
 Query Match 2.9%; Score 365; DB 21; Length 2973;  
 Best Local Similarity 19.0%; Pred. No. 2.9e-11;  
 Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;  
 QY 347 PEIKRQEQE-RFORVGQAGUSATTARSEH-----EISEIIDLSEENNEK 395  
 Db 981 psiesyeddeskfcysqypadlahkhsanhmddndgeidtpnysalkysdesqlcgr 1040  
 QY 396 QMRQLSVIPPMFQAEQRRVFINNGLMEDPMPYKDRQFMVWTDHE--KEIFKDFI 453  
 Db 1041 q-----spsqnerwarpkhileideikqsegrsqngsttppvtestddkhl 1087  
 QY 454 QHPKNFGLIASYLERKSVPCVLVLYLTCKNENKALVRRNYKRRGRNQQTARPSQEK 513  
 Db 1088 kfqphfg-----qgecvspy-----rsrgangsetnrvngnhglnqnvsgqlcqe- 1132  
 QY 514 VEEKEEDKAETEK--EEKKDEEKEDESKENTWEKIDCTAETETEREQATPRG 571  
 Db 1133 -dvyeddkptnyseryseehheerptnysikyn-eeekrhvdpidysikyadipss 1190  
 QY 572 R-----KTANSQGRRKGRITRSMTNEAAS-----AAAAAATPEPPPLPPP 614  
 Db 1191 qkqsfssksssgdsktehmssssentstpsnakrnqnlhpsaqsgrsgqpkaatck 1250  
 QY 615 PEPISTEPVEV-----SRWTEEMEVAKKGLVEHGRN-----WAAIAKM 653  
 Db 1251 vssinqetiqcyvedtpicfscsslsssaedeigcngtqeadsantliqlaeikek 1310  
 QY 654 VGTKE-----AQCKNFYNYKRRHNLNLLQHK-----QKTSR 688  
 Db 1311 igtrsaeqpvsevpavsqhprtkssrlqgsslsasesarkhavfessgskspksagatpk 1370  
 QY 689 KPREE-----RDVSOQSVASTVSAQDEDEIASNEENPEDSEAVEAVKPSE--DSPEN 740  
 Db 1371 sppehyvqetplmfsrcstsv-ssidsfseisvassvqepscsmvsgliisdlpdsppq 1429  
 QY 741 A--TSRGNTPEPAVELEPTTETAPSTPSLAVPSTKPADESVEVQVNDISIAETAQMDV 798  
 Db 1430 tmppsrsktppp---ppqtaqtkevknkaptaeakresgpkqaavnaavq-rvqvlpda 1485  
 QY 799 DQOESAAE-----GSVC-----DPPATKADSDVDVEVR-----VPEN-HASKVEGDN 840  
 Db 1486 dliihfatestpdpfscsssisalsldepfqg----dveirmpvpqndngneteseq 1541  
 QY 841 TKE--RDLDRASEKVEPRDEDLVVAQIINAQRPEPQSDNDSSATCSADEVD----- 890  
 Db 1542 pkesnengekaektidsekdl-----ddsd-----dddleileecilis 1581  
 QY 891 ---GEPERQRMFPMDSKPSLLNPTGSLVSSPLKPNPLD-----LP-----QLQHRAAVIP 938  
 Db 1582 amptkssrkakpqaatasklppvar-----kpsqlpvyklpsqrnlpqkhvstfp 1634  
 QY 939 ---PMVSC---TPCNPIPIGTPVSGYAL-----YORHIKAMHES 970  
 Db 1635 gddmrvyvcvegtplnfstatslsldtiesppnelaagevrgvgaqsgfekrdtpteg 1694  
 QY 971 AL-----LEQRQREQIDLECRSSYSPCGTSKSPNREWEVL----- 1007  
 Db 1695 rstdeagggktsvvtipeidndkkaeegdilaecinsampkshkpfvrkkmidqvqas 1754  
 QY 1008 --QPAPHQLITNLPEGVRLPTRTRPPPLI-----PS 1039  
 Db 1755 assapnkn---nqldgkkkptsvpkplpqnteytrvrknadsknnlnaervfsdnkds 1811  
 QY 1040 SKTVVASEKPSF-----IMGSGSIS-----QGTPTGYLTSHNOASTQETPK 1080  
 Db 1812 kkqnlnknskdfndklpnndrvrgsfafdsphhytlegtp--ycfsrnd----- 1860





Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

```
QY 347 PEIRKOREOE-RFORVGORGAGLSATIRSEH-----EISEIDGLSEONNEK 395
Db 981 psiesyddeskfcsyqypadlahkihsanmddndgldtpinylskysdeqlnsgr 1040
QY 396 OMROLSVIPPMMFDAEORRYKFINNMGLMEDPMKYKDRQFMNWTDHE--KEIFKDKFI 453
Db 1041 q-----spsqnerwarphkhiedelkqseqrsgstqtpyvytestdkhl 1087
QY 454 QHPKNFGLIASYLERKSVPCVLYIYITKKNENYKALVRNRYKRRGRNQOIARPSOEK 513
Db 1088 kfghfig-----qqccvapy-----rsrgangsetnrvgshngindqnvsgslcqe- 1132
QY 514 VEKEEDEKAETK-----EKKDEDEKEDKENTKEDKIDGTAEETEEREQAAPRG 571
Db 1133 -dayeddakptnyseryseeeqheerptnyskyn-eehrhvdqpidyalkyatdipss 1190
QY 572 R-----KTANSQGRKRGTRITSMTEAAAAA-----AAAAAATEEPPLPPPP 614
Db 1191 qkqsfkskssgsgsktehmssasentstpsnakrqnlhpsaqsrggqqkaatk 1250
QY 615 PEPSTPEVT-----SRTEEMEVAKKGLVEHGRN-----WAAAKM 653
Db 1251 vssingetlqtycvetpfcfscrsslssaaedeigcngttqeadsantlgliaeike 1310
QY 654 VGTKSE-----AQCKNFYFNKRRHNLNLLQOHR-----OKTSR 688
Db 1311 lgrtsaedpvyevpavqshprtkssrlqgsslsesarhkavefssgakspsksgadt 1370
QY 689 KPREE-----RDVQCESVASTVSAQEDIEDIRASNEENPEDEVAVPKSE--DSPEN 740
Db 1371 spchyyvetplmfrcstsv-saldsfesriassvqpcsgmvgslispsdlpdsppg 1429
QY 741 A--TSRCNTPAVELEPTETAPTSTPSLAVPTKPADESVETQVNDISIETAEQMDV 798
Db 1430 tmpsrsktppp---ppqtqtrevpknkaptaeekresgpkqaavnaavq-rvqvlpda 1485
QY 799 DQOHSAAE-----GSVC-----DPPATKADSVDEVR-----YPEN-HASKVBGDN 840
Db 1486 dtllhfatestpbgfscssalsaldepiqk----dvelrmpvqndngneteseq 1541
QY 841 TKE--RDLDRASEKVEPRDELVVAAQINQAORPEQSDNDSSATCSADEVD----- 890
Db 1542 pkesnenekaeaktidsekdl-----ddsd-----dddieleeciis 1581
QY 891 ---CEPERQRMFMDRPSLINTGSTLVSSPLKPNLD---LP-----QLQHRAAVIP 938
Db 1582 amptksrkakpaqtasklpppvar-----kpsqlpvyklpsqnrllpqdkhvsft 1634
QY 939 ---PMVSC---TPCNTPIGTPVSGVAL-----YORHIKAMHES 970
Db 1635 gdmpryvcvgtplnfstatisldtiespphelaagevrgaqsgefekrdtipteg 1694
QY 971 AL-----LEEQRQRQIDLECRSSTPCGTSKSPNREWEVL----- 1007
Db 1695 rstdeaggktssvtipelddnkaeegdilaecinsampgkshkpfvrkkimdvqgas 1754
QY 1008 --QAPAHQILTNLPEGVRLPTTRPPPLI-----PS 1039
Db 1755 assapnk---nqldgkkktpsvkpiqnteytrvrknadsknlnaervfsdnkds 1811
QY 1040 SKTIVASEKPSF-----IMGGSTIS-----QCTPGTYLTSHNOASVTOETPK 1080
Db 1812 kknqknnskdndklpnnedrvrgsfafdsphhyptiegtp--ycfsrnd----- 1860
QY 1081 PSYGISLGLPROQESAKSATLPIYKOEESPRSONSQPEGLLVRAQHEGVRGRTAGAIQ 1140
Db 1861 -sIssldfddvdlrsrekaelrkakenkesakvtshte--ltsnqgsa--nktqaaia 1914
QY 1141 EGSITRGTPTSKSVESIPSLRSIGTGPALPOTGIPTEALVK-----GS 1186
Db 1915 kqpinrgqkpiilqgs-----tfqsskdipdrgaatdekqlgnfaientpvcfshnss 1968
```

```
QY 1187 ISRM-----PIEDSSPEKGREAAASKGHVIEGCKSGHI----- 1219
Db 1969 Issalsdidqennnkenepiketepdpsqgpepskpaasyapksfhvedtpvcfarnsels 2028
QY 1220 -LSYDN-----IKNAREGTRSPRTAHEISLKRSESVESGNIKQOGMSRESVSPASLE 1270
Db 2029 slsidseddillqecissamnpkkpsrlkgdnekhspnmnggilgedltdlkdlqrpds 2088
QY 1271 GLICRALPRGSPHDLKERTVLSGSIMQGTTPRATTESFEDGLKYPKQIKRESPPIRAFEG 1330
Db 2089 e---hglspdsenfwdkalgagansivsslhqaaaac-----lsrqassdsdlsilks 2140
QY 1331 AITKGGPYDGIITTIKEMGRSITHEIPRODILTOESRKTPVEVVQSTRPIEGSISOGTPIKF 1390
Db 2141 gislgspf-----hltpdqeekptankprik-----pgekstlektki 2181
QY 1391 DNNS-GOSAIKHNKNSLITGP-----SKLSRGM-PPLEIVPENIKVBERGYEDVYKAGETV 1444
Db 2182 esesgikggkvykylitgkvrnselsgmkqlq---anmpsisrg----- 2227
QY 1445 RSRHTSVVSSGSPVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRCSPMNRTSD 1504
Db 2228 -----rTmh-----ipgy-----rnssatspvsykkpp 2252
QY 1505 VTIPPKNKSTNHERKSTLTPT-QRESIPAK-SPVPGVDPVWVSHSPFDPHHRGSTAGEVYWS 1562
Db 2253 lktpaskpsseqtattisprgakpsvkseispvarqtsqiggskapsrsgsr----- 2305
QY 1563 HLPTQLDPAMPFHRALDPAAYLFRQLSPTPGYPSQYQIYAMENTRQTILNDYITSQQ 1622
Db 2306 -----dstpsrpaqql----- 2320
QY 1623 MOVNLRPDVA---RGLSPREQPLGLPYPATRGIIIDLNMPTILVPHPGTSTPMPBRIT 1679
Db 2321 iqspgrnsispggrngisppnk-----lsqilprt-sspstaskssgsgkms 2365
QY 1680 YIFGTQITFFPRPNYSASMSGPHPTHLAAAASAERERERE-----KERERE 1727
Db 2366 y-----tspgqrmsqnlk--katglsknaassipresasaqngnmngngankkvels 2417
QY 1728 RIAAASDLYLRPGSEOPGRPGSHGYVRSPPSVRTQETMLQQRPSVFGTNGTSVITPL 1787
Db 2418 rmstskss-----gsesd-----rserpvlrvgtfikeaps----- 2449
QY 1788 DPTAQLRIMPLPAGGPSISQGLPASRYNTAADALAAALVDAASAPOMDVSKTKESKHEAA 1847
Db 2450 -ptlr-rkieesafesis---psrpsatrsqatcpvispdmnsls-thssvgaqg 2503
QY 1848 --RLEENLRSRSAVVSQOQLEQKTELEVERKRSVQCLYTSSAFFSGKPPQHSSVVVYSBAGK 1905
Db 2504 wrklppnl---sptieyndgrpakrhdiar-----shsesparlpinrgtwtkrehs 2553
QY 1906 DKGPPPKSRVEELRTRGKTTITAAAN-----FIDVITRQIASDKDARERSQSSDS 1957
Db 2554 hssslp-rvstwrtrgtssslsassesekaksekdekhvnsigtkqskencvsaagt 2611
QY 1958 SSSLSSHRYE-TPSDATEVISPASSAPPQOEKLOTYQPEVVKANQAEENDPTROYEGPLHH 2016
Db 2612 wrkikeneftsptustqtsvsgatngaesktllymqapavsk---tedvvrriedcplnn 2668
QY 2017 YRPQOESPSQQQLPP-----SSQAEQMGQVPRTHRLITLADHIC 2056
Db 2669 pr---sgrsptgntppvidsvsekanpniksdnkqakqngvngsvpm--rtvglenrln 2723
QY 2057 QIITQDQFARNQVSSQTP-QQOPTSTFQNSPALSIV-TPVRTKTSNRYSPESQAOQSVHHQR 2114
Db 2724 sfliqvdapdkgteikpgqnpvpsvtnessliverpfssssskkbspsgtva----- 2778
QY 2115 PGRSVSPENLVDRSGRPGKSPERSHVSSEPEPISPPQVPPVH--EKQDS 2164
Db 2779 --arvtpfn-----npsprkssadstsrpsqip-tpvnnntkkrds 2818
```

RESULT 26  
AAW11922  
ID AAW11922 standard; protein; 2843 AA.  
XX  
AC AAW11922;  
XX  
XX 07-MAY-1997 (first entry)  
XX  
XX Adenomatous polyposis coli protein.  
DE  
XX Adenomatous polyposis coli; APC; human; antibody; mutant detection.  
KW  
XX Homo sapiens.  
XX  
XX JP06347459-A.  
XX  
XX 22-DEC-1994.  
XX  
XX 07-JUN-1993; 93JP-0136102.  
XX  
XX 07-JUN-1993; 93JP-0136102.  
XX  
XX (SANK-) SANKO JUNYAKU CO LTD.  
XX  
XX WPI; 1995-070957/10.  
XX  
XX Detection of abnormal human APC (adenomatous polyposis coli) gene -  
PT using antibodies directed against the APC gene product  
XX  
XX Claim 1; Page 7-13; 13pp; Japanese.  
XX  
XX This sequence represents the wild type human adenomatous polyposis coli  
CC (APC) protein. This sequence, and the APC peptide fragments represented  
CC by AAW11922-W1929 can be used in the method of the invention. The  
CC method of the invention is for the detection of abnormal APC, using  
CC antibodies directed against this sequence, or one of the partial peptide  
CC sequences.  
XX  
XX Sequence 2843 AA;  
XX  
Query Match 2.9%; Score 364; DB 16; Length 2843;  
Best Local Similarity 19.0%; Pred. No. 3.1e-11;  
Matches 409; Conservative 290; Mismatches 803; Indels 652; Gaps 95;  
QY 347 PEIRKQEQE-RQRYGQRCAGLSATARSEH-----EISEIIDLSEQENNEK 395  
DB 981 pslesyeddeskfcsygyqypadiahkhihsanhmddngeldtpnyslkysdeqlnsgr 1040  
QY 396 QMROLSVIPPMMFDAQRRVKFINMGLMEDPMKVYKDRQFMNVWTDHE--KEIFKDKFI 453  
DB 1041 q-----spqnerwarphkhiiedelkqseqrgsttyppytestdckhl 1087  
QY 454 QHPKNFGLIASYLRKSVPCVLYYYTKKNENYKALVRRNYGKRRGNQOIARPSQEEK 513  
DB 1088 kfqphfg-----qqecvspy-----rsrgangsetnrvgshngingnvsqslqce- 1132  
QY 514 VEERKEKAEKTEKK--EEKKDEEKDEKEDKENTKDKIDGTAEETEERQAPRG 571  
DB 1133 -dyeddqkptnysyreeeqheerptnysikyn-eekrhvdqpidyskyatdlps 1190  
QY 572 R-----KTANSQRRKGRITRSMTEAAAAA-----AAAAATEEPPLPPP 614  
DB 1191 qkqsfksksgsgskstehmssasntstpsnakqnlhpsaqsrggqkaatk 1250  
QY 615 PEPISTEPVEV-----SRWTEEMEVAKGLVEHGRN-----WAAIAK 653  
DB 1251 vssingetigtvcvedtpicfscsslsissaeideicnqtqeadsantlqlaeikek 1310  
QY 654 VGTKSE-----AQCNFYFNFKRRNLDNLLQOHK-----OKTSR 688  
DB 1311 igrtsaedpvsavpavsqhprtkssrlqgsslsesarhkavsfsgakspksqsgatpk 1370

QY 689 KPREE-----RDVSQCESVASTVSAQEDIEDIASNEENPEDESEVAVKPSE--DSPEN 740  
DB 1371 sppehyvqetplmfsrctsv-ssldsfeisaiassvsepccsmgmvsliispsdipdsppq 1429  
QY 741 A--TSRGNTPEPAVELEPTTETAPSTSPSLAVPSTKPAEDESIVETQVNDISAEAEQMDV 798  
DB 1430 tmppsrsktppp---ppqatkrvnpknkaptakresgpkqaavnaavq-rvqvlpda 1485  
QY 799 DQOEHSAAEE-----GSVC-----DPPPATKADSDVEVR-----VPEN-HASKVGDN 840  
DB 1486 dtllhfatestpdcgfcscsslsalsldpfiqk----dvelrmpvqndngneteseq 1541  
QY 841 TKE--RDLDRASEKVEPRDEDLVVAQIQNAQRPEQSDNDSSATCSADEVD----- 890  
DB 1542 pkesnenqeakektidsekdll-----ddsd-----dddielleeciis 1581  
QY 891 ---GEPERQRMFPMDSKPSSLNPTGSLVSPKPNPLD---LP-----QLQHRAAVIP 938  
DB 1582 amptkssrkakkaqaasklpppvar-----kpsqlpvykllpsqnrlqpqkhvstfp 1634  
QY 939 ----PMVSC---TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
DB 1635 gddmrvycvegtpinfstatsldtiesppnelaagevriggaqgefekrdtpteg 1694  
QY 971 AL-----LEEQRQRQEQIDLECRSSSPCGTSKSPNREWEVL----- 1007  
DB 1695 rstdeaggtssvtipelddnkaeegdliaecinsampkghkpfvrvkkindqvqas 1754  
QY 1008 --QPAPHOLITNPEGVRLPTTRTPRPPLI-----IMGGSIS-----QCTPGTYLTSHNQASTQTTPK 1039  
DB 1755 assapnk---ngldgkkkxptspvpipqnteytrvrknadsknnlnaervfsdnkds 1811  
QY 1040 SKRTVASEKPSF-----IMGGSIS-----QCTPGTYLTSHNQASTQTTPK 1080  
DB 1812 kgnlnknsdkfndklnpnnedrvrgsfafdsphhytplegt-ycfsrnd----- 1860  
QY 1081 PSVGSISLGLPRQOESAKSATLPYIKOEFSRSONSQEGLLVRAQHEGVGRGTAGAIQ 1140  
DB 1861 -sissidfdddvdlsrekaelrtakeneseakvtshte--ltsnqsa--nktgaia 1914  
QY 1141 EGSITRGTPTSKISVESIPSLRGSITOGTTPALPOTGTPTEALVK-----GS 1186  
DB 1915 kqplnrqgkpliqkgs-----tfqsskdiipdrgaatdekignfaientpvcfshns 1968  
QY 1187 ISRM-----PIEDSSPEKGREAAASKGHVYIEGKSCHI----- 1219  
DB 1969 lsalsdidqannkenepiketepdpdsqgepskpaasgypaksfhvedtpvcfsrnsals 2028  
QY 1220 -LSYDN-----IKNAREGTRSPRTAHEISLKRYSVESVEGNIKQGMRESVPSAPLE 1270  
DB 2029 slsidseddillqecissampkkkpsrkgdnekhspnmnggilgedltldldkdiqrps 2088  
QY 1271 GLICRALPRGSPHSDLKERTIVLSGIMOGTPRATWESFEDGLKYPKQIKRESPIRAFEG 1330  
DB 2089 e---hgispdsenfwdkaigegansivsslhqaaaaac-----lsrqassdsilsilks 2140  
QY 1331 AITKGPYDGIITTIKEMGRSIEHETPRODILTQESRKTPVEVQSTRPIEGSISQGTPIKF 1390  
DB 2141 gislgspf-----hitpdqeeqkftsnkgprilk-----pgekstlektki 2181  
QY 1391 DNNS-GOSAIKHNKVSILITGP-----SKLSRGM-PPLEIVPENIKVBERGYEDVKAETV 1444  
DB 2182 eseskgigkgkkykslitgkvrnsneisgmkqplq---anmpsifrg----- 2227  
QY 1445 RSRTSVSVSGSPVLSLTLHEAPKAQLSPGIYDDTSARRTPVSVQNTMSRGSPPMNTSD 1504  
DB 2228 -----rtmih-----ipgv-----rnsststspvskkgpp 2252  
QY 1505 VTTPPNKSTNHERKSTLTPT-QRESIPAK-SPVPGVDVPSVSHSPDFPHHGRSGAGEVWS 1562  
DB 2253 lktpaskspsegqtattsprgakpsvkselspvarqtsqigsgskapsrsgr----- 2305

QY 1563 HLPQLDPAMPFHRALDPAALAAALFQRLSPTPGYPSQYQLYAMENTROTILNDYITSQQ 1622  
Db 2306 -----dstpsraqpl-----snp 2320  
QY 1623 MOVNLRPDVA---RGLSPREQPLGLPYPAYRGIIDLNTMPPTILVHPGCTSTPPMDRIT 1679  
Db 2321 iqspgrnsipgrngisppnk-----lsqprt-ssptastcksgsgkms 2365  
QY 1680 YIPGQTITFFPRPNYSASPGPHTHLAAAASABERERERE-----KERERE 1727  
Db 2366 y-----tsgrmqsqnlt--kqtglsknassipresasakngnmgnganakkvels 2417  
QY 1728 RIAAASDLVLRPGSEOPGRPGSHGYVRSPSPSVRTQETMLQRPSPVFGTNGTSVITPL 1787  
Db 2418 rmsstks-----gsesd-----rserpvlrvstfikeaps----- 2449  
QY 1788 DPTAQLRIMPLPAGGPSISQGLPASRYNTAADALAAALVDAASAPQMDVSKTESKHEAA 1847  
Db 2450 -ptlr-kliesasafesis---psrpsptrsqatqtpvislpdmsis-thsvagag 2503  
QY 1848 --RLEENLRSRSAVSEQQOLEQKTLVEKRKRSVQCLYTSSAFPPSGKQPHSSVVYSEAGK 1905  
Db 2504 wrklppnl---sptieyngdrpkrhdiar-----shsesprlpinrsgtwkrehs 2553  
QY 1906 DKGPSPKSRVEELRKGTTITAN-----FIDVITRQIASDKDARERQSQSDS 1957  
Db 2554 hssslp--rvstwrirtgsssilssasesseksaksedekhvnsisgckskengvsaagt 2611  
QY 1958 SSSLSHRYE--TPSDALEVISPASSPAPQEKLOTYOPEVVKANQANENDPTROYEGPLHH 2016  
Db 2612 wrlikenefstnqtstqvsagatngaesktllyqmapavsk-----tedvvrriedcpinn 2668  
QY 2017 YRPOESPSPOOQLPP-----SSQAEGMGQVPRTHRLITLADHIC 2056  
Db 2669 pr---srsptgntppvidsvsekanpnikdskdnqakngvngsvpm--rtvgleln 2723  
QY 2057 QIITQDFARNQVSSQTP-QOPPTSTFQNSPVALVS-TPVTRKTSNRYSPESQAQSVHHQR 2114  
Db 2724 sfidvdpdkgkteikpgqnpvpvsetnessiverfssssskhspsgtva----- 2778  
QY 2115 PGSRVSPENLVDRSGRPGKSPRSHVSEPEYPIPPQVPV-----HEKQDS 2164  
Db 2779 --arvtgfn-----npsprks--sadsys-arpsqiptvnnntkrds 2818  
RESULT 27  
ID AAY57453  
AC AAY57453  
XX AAY57453;  
XX 22-FEB-2000 (first entry)  
XX Human transcriptional regulatory factor SEQ ID NO:10.  
XX DE  
XX KW Human; transcriptional regulatory factor; TCOA1; BLAST detection;  
XX KW bromo-domain; cell proliferation; cancer.  
XX OS Homo sapiens.  
XX PN WO9957143-A1;  
XX 11-NOV-1999.  
XX PF 30-APR-1999; 99WO-JP02340.  
XX PR 30-APR-1998; 98JP-0137631.  
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
XX PI Jones MH;  
XX WPI; 2000-052940/04.  
DR

DR N-PSDB; AA239033.  
XX Transcriptional regulatory factor containing a bromo domain and gene  
PT TCOA1 encoding it  
XX Claim 1; Page 139-151; 154pp; Japanese.  
XX The present sequence represents a human transcriptional regulatory factor  
CC containing a bromo domain. The factor interacts with proteins involved  
CC in the chromatin-mediated transcription regulation mechanism. It binds  
CC to hSNF2H, hSNF2L and NCOA-62/Skip. It can be used for screening  
CC compounds binding to it and acting as agonists or antagonists, which  
CC are potentially useful for the treatment and prevention of cancer and  
CC other cell proliferation disorders.  
XX Sequence 2781 AA;  
SQ  
Query Match 2.8%; Score 356.5; DB 21; Length 2781;  
Best Local Similarity 17.6%; Pred. No. 7.7e-11;  
Matches 449; Conservative 354; Mismatches 889; Indels 865; Gaps 108;  
QY 97 SPVDHDSLESKRPRLEQVSDSHFQVSAAVLPLVHPPLPEGLRASADAKKDPAFGGKHEAP 156  
Db 270 savyh--lecvkplleevpedewq-----cevcvahkvpvgvtdcvaeiqnky 316  
QY 157 SSPTSGQPCG-----DDQNASPSKL-----SKEELIQMDRVDRDRIA 193  
Db 317 ---irhepigydrsrkywflnrrliieedtenenekkiwyystkvqlaelidcldky 373  
QY 194 KVEOOILKLLKKQOOLEEAAKPEPEKVPSPVPEQKHSRIVQIYYDENRKK-----AE 248  
Db 374 eae-----lckileemreei-----hrhmditedltnkargnsksflaa 412  
QY 249 EAHKIFEGL-CPKVELPLYNQPSDTKYVHENIKTNQVWRKKLILFFRRNRHARKQREOKI 307  
Db 413 aneilesirakkdidnvspeete-----kdknetendskda- 451  
QY 308 CORYDQLMEAWKVKVDRIENPNRKAESKT-----REYVEKQFPEIRKQREQQERFORVG 363  
Db 452 -----eknefeedsglskdsddktpdddppegksevdfksekngelsesp 500  
QY 364 QRGAGLSAT-----TARSEHEISEIIDGLSE--QENNEKQMRQL 400  
Db 501 -agkagsgstriitrlrnpdsklsqksqvaahankifkegkevlvvnsgseisrl 559  
QY 401 SVIPPMFEDAEORRVKFTINMGL---MEDPMKYVKDQFMNVT-----DHEK-EI 447  
Db 560 stkkevimkg-----ninnyfklgqegkyrvyhnqystnsfalnkhqhrdhdkrrh 611  
QY 448 FKDKFIQHPK-NFGLIASYLERKSVPCVLYYLTKKENY-KALVRRNYGKRGR----- 501  
Db 612 lahfclitpagefwkngsvhskvltistrlitqlennipssfilhpnwshranwika 671  
QY 502 NQIARPSQ-----EEKVEEKDEKAEKTEKKE 529  
Db 672 vqmskprefalalailailecavkvpmplwreflightrlhrmtsiereekekvkkek 731  
QY 530 EEKDEEEKDEKESKENTKEKIDGTAEETEEREQATPRGRKTANSQGRKGRITRSM 589  
Db 732 eee-----etmqat----- 741  
QY 590 TNEAAAASAAAAATEPPPLPPPPPEIPSTPEVETSQTE--EEMEYAKGLVEHGRNW 647  
Db 742 -----wvkytfvkhqwkqgeeyrvtyg-----gsw 771  
QY 648 AA-----TAKMVGTKSEACKNFYFNKR-----RHNLN-LNLOQHKQKTSRKP- 692  
Db 772 iskthvyfvpklpg-----ntnvnrykslegtknmdenmdesdkrcspskkik 823  
QY 693 -----ERD-----VSOCESVASTVSAQEDIED--ASNEEENP----- 723  
Db 824 iepdsekdevkgsdaakgadqnmndiskitekkgdqvkelldsdskpckkeepnevddm 883

QY 724 -----EDSEYAVKPS-----DSPENATSRGNTPEAVE----- 752  
Db 884 kteshvnqessqdvvnvsegflrltsykkktskldglllerikqftleekqrleki 943  
QY 753 -----LEPTTAPSTPSLAVPTKPAEDSVETQVNDISIAETAQMDVDQOEH----- 803  
Db 944 klegikgigtstnsenkl---sespvitkakeqcsdmrgeqsnnandqpeditliq 1000  
QY 804 -SAGEGV---CDPPATK---ADSV---DVEVRVPENHASK---VEGDNTKERDLDRAS 850  
Db 1001 csqsdssvlrmsdpshtnklpykdrldvdiirspcktpkqnsiend-leeekvsdlas 1059  
QY 851 EKVEP-----RDEDLVVAQOINAOQRE-P 873  
Db 1060 rgqgetsktgnaffiddsklasaddigtlicknkkpligeesdtivssksalhesvp 1119  
QY 874 QSDNDSSAT-----CSADEDGEPERQRMFMD-SKPSLLNPTGSLVSSP 919  
Db 1120 kstndratpismrdfeqkigcdsesnstlienssdvtsiqdsseedmivqnsesieq 1179  
QY 920 LKPNPLDLQLOHRAAVIPPMVSTPC--NIPi-GTPVSGYALYORHIKAMHESALLIEQ 976  
Db 1180 ftrreqdvevlepkcelvsgestgncedrllpvkgteang-----kpsqgkkleer 1231  
QY 977 RQROEQIDLECRSSTSPCGTSKSPNREWEV----- 1006  
Db 1232 pvnkcsdgiiknttd---kknenresekgqstftfgingkdkkylkgeclkeis 1288  
QY 1007 -----LOPAPHQILTNLPEG-----VRLPTRTRPPPLIPSKTTVASEKPSFI 1052  
Db 1289 esrvvsgnvepkninkilpendiksitvkesalrfingdvimedfnerussetkshl 1348  
QY 1053 MGSISQCTPGTYLTHNQASVYQETPKPSVGSISLGLPRQ-----QESAKS 1099  
Db 1349 issdae---gnyrdsetlpstkesdttqtpsascpesnvqvedmeletsevk 1405  
QY 1100 ATLPIYIKOEFS-----PRSQNSOPEG-----LLVRAQHEGVVR 1133  
Db 1406 tsspitseesnlndfidenglpinknenvngeskrktvitevtmtstvatetsktvik 1465  
QY 1134 CTAGAIQ-----BGSITRGTPTSKISVESIPSLURGI-----TQGT 1169  
Db 1466 vekgdkqtvvastencakstvt---ttttvtklstpsgtsgsvdiisvkeqsktvtvtvt 1523  
QY 1170 PALPTG-----IPTEALVKGISRMPIEDSSPEKREAAKSGHVIYEGKSG--- 1217  
Db 1524 dsltttgglvtmtskeystrdkvklmkf--srpkktirsgtalpsyrkftvktkski 1581  
QY 1218 HILSYDNIKN-AREG-----TRSPRTAHEISLKRYSVESVEGNIQ 1256  
Db 1582 fvlpnddikkliarkgirevpyfynakpalidiwypsprptfgitwryrlqtkvkslagv 1641  
QY 1257 GMSHRESVPVAPLEGICRALPRGSPHSDLKERTVLSGSIHQGTPTATSFEDGLKVPK 1316  
Db 1642 slmlrllwaslrwdmaakvppgg-----gstr--tetseteittte 1681  
QY 1317 QIKRE--SPPIRAEGATKPKVDGITTIKEMGSIHEIPRODILLQESKRTPEVQST 1374  
Db 1682 likrrdvpgyrlrfeycirk-----ilcpig--vpctketptqrkqlrssairpk 1731  
QY 1375 PIITEGSIQGTPTIKFNNNSQSAIKHNKSLITGPKSLSRGMPPLEIVPNIKV---V 1430  
Db 1732 rp-----etpkq-----tgpyietwaeelweirafaerv 1765  
QY 1431 ERKYEDV--KAGETVRRHRTSVV-----SSGPSVLRTLHEAPKAQLSPGIYDTSARRT 1484  
Db 1766 ekekaqaveqqakkrleqktpviatsttsptssttstispaqvmvapisgsvttgktm 1825  
QY 1485 PVSQNTMSRSGSPMNRSTSDVTPPNKSTNHERKSTLTPTQRES--IPAKSPVPGVDPVV 1542  
Db 1826 vi-----ttkvgspp-----atvtfqgnknfhqtfatwkgqsgnsqvvgqkvligi----- 1872

QY 1543 SHSPFDPHHRGSTAGEYVWVSHLPTOLDPAMPFHRALDPAAAYLFOQLSPTCPYQSOY 1602  
Db 1873 -----ipststgtsqftsfqprtattvirtpntsgsggtsnsq 1911  
QY 1603 LYAMENTROTILNDYITSQOMVNLRPDVARGLSPREQ-PLGLPYPATRGIIDITNMPPT 1661  
Db 1912 v-----itgpg---irpgmtvirtpqlqstlg-----kali-----rt 1941  
QY 1662 ILVPHPGCTSTPPMDRITYIPGTOITFPPRPNYSASMSPGHPHAAAASAEERERERE 1721  
Db 1942 pvmvqpg---apqgvmtqilrgqpvstavsapntsvstpg----- 1978  
QY 1722 KERERERIAAASDLYLRPGSEQPRGSGHYVRSPSVRTOETMLQORPSVFGTNGT 1781  
Db 1979 ---qksltatsstnsiqsasqprrp-qggqv-----ltmaqltqlcqhggn 2023  
QY 1782 SVITPL-----DPAQRIIMP-----LPAGPSISQ--GLPASRYNTAADALAAALVDAAS 1830  
Db 2024 qglitvliqggggttqglilipqgvvlpqpgqilmqampngtvqrflftplattattas 2083  
QY 1831 APQMDVSKTKESKHEAARLEENLRSRAAYSEQQOLEQ-KTLEVEKRS---VQCLYTSSA 1886  
Db 2084 tttttvtaagtge-----grqsklspgmqvhdktlppagssvvpakagpqa 2134  
QY 1887 FPSGKPOPHSHSVVYSEAGKDKGPPPKRYEELRTRG--KTTITAAFNIDVITRQIASD 1944  
Db 2135 qpsarpapqtq-----pqspaqpevtqpevtqttvsshv-----p 2171  
QY 1945 KDAREGQSQSDSSSSLSHRYETPDSAEIVSPASSPAPPQEKLOIYQP-EVVKANQAE 2003  
Db 2172 seaqthagskspqvaasqg---pqsnvqgspvrvqgspqtrirpstpsqlspgqsg 2227  
QY 2004 NDPTRQEGPLHHYR---PQESPSQQQLPPSSQAEGMGQVPRHRLITLADHICQII 2059  
Db 2228 vqttsgpikiphtsliqipesqgqgqpsqvtqstqlsssq-----tl----- 2271  
QY 2060 TQDFARNOVSQTPQOPPTSTFQNSALVSTPVRTKTSNNRSPESQ-AQSVHHQRPGSR 2118  
Db 2272 -----nqvsvsspsrpqlqilqppqvaiavpqlqgqvqlsqgsvvvaqiaqgsg-- 2323  
QY 2119 VSPENLVDKSRGSPGKSPERSHVSSEPEPISPQVPVYVHEKODSLLLSQSGAEPAEQ 2178  
Db 2324 -vpqgi---klqilqigssavqthqgnvvtvqaasvqeqqlrvqqlrdqgqkqkq 2378  
QY 2179 RNDARPGSISYLPSTFKLENTSPMVKSQKQEIFRK 2215  
Db 2379 qie-----ikreht--lqasngseliql 2399  
RESULT 28  
AAB50362  
ID AAB50362 standard; protein; 3118 AA.  
XX AAB50362;  
AC AC  
XX AC  
DT 12-MAR-2001 (first entry)  
XX Human SRCAP.  
DE Human.  
XX Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;  
KW CAMP regulatory element; CREB binding protein; CBP; ATPase;  
KW transcription activation; DEAD box RNA dependent helicase;  
KW adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.  
XX Homo sapiens.  
OS Homo sapiens.  
XX WO200073467-A1.  
PN PN  
XX 07-DEC-2000.  
PD PD  
XX 25-MAY-2000; 2000WO-US14719.  
PF PF  
XX 27-MAY-1999; 99US-0136620.  
PR PR







Db 2434 qiqsqvvaqigaq 2446

## RESULT 30

AAB50363  
ID AAB50363 standard; protein; 2972 AA.

XX AAB50363;

AC 12-MAR-2001 (first entry)

DE Human SRCAP.

KW Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;  
KW CAMP regulatory element; CREB binding protein; CBP; ATPase;  
KW transcription activation; DEAD box RNA dependent helicase;  
KW adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.

XX Homo sapiens.

OS WO200073467-A1.

PN 07-DEC-2000.

XX 25-MAY-2000; 2000WO-US14719.

XX 27-MAY-1999; 99US-0136620.

PR 25-MAY-2000; 2000US-0579181.

XX (UYSL-) UNIV SAINT LOUIS.

PA Chrivia J, Yaciuk P;

PI WPI; 2001-061545/07.

XX N-PSDB; AAC89860.

DR Snf2 related CAMP regulatory element (CREB) binding protein (CBP)  
XX activator protein, capable of co-activating CREB binding protein,  
PT useful for modulating transcription and for affecting viral infection -

XX Claim 1; Page 86-94; 103pp; English.

CC The present sequence is an Snf2 related CREB (cAMP regulatory element)  
CC binding protein (CBP) activator protein (SRCAP) polypeptide. It has  
CC ATPase activity and is capable of activating transcription. SRCAP  
CC polypeptides are useful for activating transcription in a cell, for  
CC enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated  
CC activation of transcription in a cell, for treating a patient having a  
CC disease involving a function such as insufficient transcription of a  
CC gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent  
CC helicase, adenoviral DBP protein, beta-actin or a nuclear receptor  
CC affected by SRCAP protein. Compounds that modulate SRCAP function, such  
CC as antibodies, antisense molecules, polynucleotides or ribozymes, are  
CC useful for treating diseases mediated by SRCAP-activated transcription,  
CC for example, infection by adenovirus, hepatitis C virus, human  
CC immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or  
CC hepatitis B virus.

XX Sequence 2972 AA;

Query Match 2.6%; Score 329; DB 22; Length 2972;

Best Local Similarity 18.6%; Pred. No. 2.7e-09;

Matches 472; Conservative 300; Mismatches 866; Indels 904; Gaps 120;

Qy 432 KD-RQPMNVWDHKEI-FKDKFQHPKFNGLIASYLERKSVPCVLYY-VLTKKNNYK 488

Db 3 kdvrfq---wnvekvvqfkgq-----srleekrkkaldhldfivgqteky 47

Qy 489 ALVRNRYKRRGNQOIAARPQOEKVEKE---EDKAETKEKEEKKDEE-----EKD 539

Db 48 dli-----sgslngqitssakgsspcslgsssaasspppsarldddedgdfqpdq 97

Qy 540 EKEDSKENTKEDKIDGTAEETEEREQATPRGRKTANSQGRKRG-----RITRSM----- 589  
Db 98 eeeddeetieeegqegndaaqrreieell-----iregelpleellrslppqll 147  
Qy 590 -----TNEAASASAAAAATEEPPLP-----PPPEPIS----- 619  
Db 148 egpspsqtspsshdstrdgpgegaeeeeppqvylelkpppsavtqknkpwphdpdedest 207  
Qy 620 -----TEPVETSRWTEEMEVAKKGLVEHGRNWAIAIAKMGVTKSEAQCCKNFYFKRRHN 674  
Db 208 aneeaeedeetiaaeqle-----gevhamelselar-----egels 246  
Qy 675 LDNLLOOH-----KQKTSRKPREERDV-----SQCB-----SVASTVSAQDEDEASNEEN 722  
Db 247 meellqyagayagsgssdededevdandsdcepegvaeepppdssqsdsvdr 306  
Qy 723 PEDSEVAKPSSEDPENATSRGNTPEAVELEPTTETAPSTSPSLAVPSTPAEDSVET 782  
Db 307 sedeedeheeeetsgsaseseesedaq-----sqsqadee 346  
Qy 783 QVNDISIAETAQMDVDQOEHSAGEGVCDDPPAT---KADSDV----- 824  
Db 347 eedddfgveyllardeeqseadagsgpp-tbgpttlgpkkeitdiaaaeslqpkgytla 405  
Qy 825 --EVRVP-----ENHASKVEGDNTKERDLDR-----AS 850  
Db 406 ttqvktpllllrgqlreyghigldwlvmyekkingilademlgkigtisillahlac 465  
Qy 851 EKVEPRDEDLVV-----AQOINARQPEPQSDNDSSATCS 884  
Db 466 ekgnwgphllivptsvmlnwemelmkrwcpkfltyygaqkerklkrtgwktpnafhvc 525  
Qy 885 AD-----ED-----VDGEPERQRMFMDSKPSLLNPTGS---ILVSSPLKPNP 924  
Db 526 tsyxlviqdhqafrrknwrylildeaguknfksqrwslfnfnsrrllltgtplqsl 585  
Qy 925 LDLQLOHRAAVIPPMVWVSCPCNI-----PIGTVPVSGYALYQRH-IKAMHE----- 969  
Db 586 melwslmh---flmhpvfqshrefkwfsnpltmiegsqeqneglvxrlhkvirpflir 642  
Qy 970 --SALLEEQRQREQIDLECRSS----- 990  
Db 643 rkvvdveqkmpkkyehvircslskrqrclydmdfmaqtktketlatghfmsvinllmqrlk 702  
Qy 991 -----TSP----- 993  
Db 703 vcnhpnlfdrprvtsfptpgicfstaslvratdvhplqlridmgrfdlliglegrvsye 762  
Qy 994 -----CGTSKSP-----NREWEVLQAPAPHQ-----LITNLPE 1020  
Db 763 adtflprhrisrrvllavatadppprpkpvkmkvn---mlqvpkqegrtvvvnnpr 819  
Qy 1021 GVRLPTRTRTPRP-PLIPSSKTTTVAESKPSFIMGSGISQGTPTGYTLTSHNOAS-----YT 1075  
Db 820 aplgqv---pvrppgpelisaqpt-----gpvpqvlpsalmsvaspaggplipa 866  
Qy 1076 QETPKPSVGSISLGLPROESAKSATLPYIKQEEFSPRSQNSQPEGLLVRAQHEGVVGT 1135  
Db 867 srppgpvl-----lpllpqn---sgslpqn-----lpspl-----gvisgt 899  
Qy 1136 AGAIQEGSITRGTPTSKISVESIP-----SLRGSIT-----QGTPALPQTGIPTEALVKG 1185  
Db 900 srpbtptlsiktpapvrispappgssslklptvpgytfppaaattsttta--ta 957  
Qy 1186 SISRMPTEDSSPEK-----GREEAASGHVIEYEGSKSHILSYDNKIKNAREGTRSPRAHE 1240  
Db 958 ttavpaptapqrlilspdmqarlpsegev---sigqlaslaqrpvanaqgskpltf-- 1012  
Qy 1241 ISLRSVESVEGNKQGMMSRSPVSPAPLEGLICRALPRGSPHDLKERTVLSGS----- 1295  
Db 1013 -----qiggn-----kiltgavqrqlavgqprplqmpmnmvntgvyki 1052



|    |      |              |                     |                        |                                         |                        |                           |      |
|----|------|--------------|---------------------|------------------------|-----------------------------------------|------------------------|---------------------------|------|
| Qy | 1296 | IMQGT        | PRATTES             | FEDGL-KYPQIKRES        | PIRAFEG-----ATKKG-----                  | -----p                 | 1333                      |      |
|    |      | :::          |                     |                        |                                         | :::                    |                           |      |
| Db | 1053 | vvqrqpr----- | dgltvpv             | lapaprpsssgl           | pavlnprptltgltgltarap                   |                        | 1305                      |      |
| Qy | 1338 | YDGT         | TIKEMGR             | SIHEIPRODIL            | TOESRKTPEWOSTRPI--TEGSI                 | SQGTPTKFDNNS           | 1394                      |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1106 | mptptlv      | rplkvhs-papev       | saasapgaapltis         | splhvssipgpassmp                        | -----p                 | 1159                      |      |
| Qy | 1395 | GQSA         | IKHNKSL             | TIGPSKLS               | RGMPLEIVPENIKVYVERGKEDYKAGETVR-SRHTSVVS | 1453                   |                           |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1160 | nsgpl        | asvstsvp--lssslp-   | isvpttlp               | -----pasapltip                          | asapltvsa              | 1208                      |      |
| Qy | 1454 | SGS          | VLRS-TLHE           | APRAQILSPGIYDDT        | SARRTPVSYQNTMRGSPMMNR-TSDVTIPNK         | 1511                   |                           |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1209 | sgpallt      | svtvp               | llavvpaag-----         | ppslq-psgaspasaltlglatapsl              | 1256                   |                           |      |
| Qy | 1512 | STN          | HERKSTUTPTQRESI-    | PAKSPVPGVDPVVS-----    | hspfd-----                              | 1548                   |                           |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1257 | s-----       | sqtpbph             | lllaptashvpg           | lnstvapacsvlp                           | vaasalapfpaapn         | 1310                      |      |
| Qy | 1549 | -----        | PHHR                | STAGEVYWSHLPTQLDP-AMPF | HRALDPAAAYLFQRQ-LSPTPG-----             | 1596                   |                           |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1311 | asillap      | asasqa-----         | latlap                 | maapctailaps                            | papplpvlaps            | 1363                      |      |
| Qy | 1597 | --           | YPSQ                | LYAMENTROTILNDYI       | YSQOVNLRPDVARGLSPREOPLGLPYPATRGIIID     | 1654                   |                           |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1364 | assqt        | pvpvmap             | sptgtslasaspv          | paptvplapsstqtmalp                      | pvppslp                | 1423                      |      |
| Qy | 1655 | LT-NMPT      | ILVPHPG-----        | GTSTP-----             | PMDRIYIPGTOITPPRPYNSASMSGP----          | 1701                   |                           |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1424 | lap          | apltlggssps         | qtslgltgnpgpft         | qtsltpasslv-ptpa                        | qtslapppl              | 1481                      |      |
| Qy | 1702 | HP           | THAAA               | SAERERERERERERERIA     | AAASSDLYLRPGSEQPGRGSHGVYRSPSPSV         | 1761                   |                           |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1482 | gptq         | tslapapp-----       | -----                  | lapaspv                                 | gap-ahltl              | 1516                      |      |
| Qy | 1762 | RTQET        | MLQORPSVFQGTNGT     | SVITPLD-PTAQL          | RTMLPAPGSPISQGL-----                    | PASRYNTA               | 1817                      |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1517 | -----        | sasll               | pasvqtltslp            | apvptlgpaaqtlal                         | lapastgpa              | 1558                      |      |
| Qy | 1818 | ADALA        | ALVDA--AASAPQ       | MDVSKTESKHEAARLEEN     | LRSSAAVSEQOOLEQKTXLEV                   | 1875                   |                           |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1559 | sqasslv      | saasgaapl           | pvtmvr                 | lrvyskde-----                           | pdltlrgppsp-----       | 1601                      |      |
| Qy | 1876 | RSVQ         | LYTSAPSGKGPQPHSS    | VVYSEAGKDGKPPKSYE----- | EELTRGKT-----                           | 1925                   |                           |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1602 | -----        | pstats              | ggpr-----              | iqppppr                                 | spfyldsleekrkqrserleri | 1645                      |      |
| Qy | 1926 | -----        | TTAN                | FTD-VIITQ              | TASDKDAREG-----                         | SQSDSSSSLSH            | 1964                      |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1646 | fqlse        | ahgalapvyg          | tevdftclbpq            | vaspigrpspsphtfwt                       | yeaahravlfppq          | 1705                      |      |
| Qy | 1965 | RYETP        | SDAIE---VISP        | ASSPAP-----            | POEKIQTQPEVVKRQANENDPTQYSGPL            | 2014                   |                           |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1706 | rlql         | seilte              | rflfvmppve             | appls                                   | lhachppvp              | lapqaafceqlaselwpr-----pl | 1762 |
| Qy | 2015 | HHYR         | POQSPSPQOQLUPP----- | SSQAEG                 | QMGVPRTHLI-----                         | TLADH                  | 2054                      |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1763 | hrivc        | mrqtqfd             | rlrlqydc               | klqtlavllrq                             | kaeg-----              | hrvlfqtmrldv              | 1815 |
| Qy | 2055 | ICQI         | ITQD---FARN         | VSSQTQQPQPTSTFQNS      | ALYSTVPVTKT-----                        | 2097                   |                           |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1816 | legfl        | thghlylrl           | dgstrve                | qrgalmernadkr                           | icfllstrsgvgv          | nlgtadv                   | 1875 |
| Qy | 2098 | --           | SNRYS               | PESQAQS-----           | VH-HQRP                                 | GSVSPENLVKSRGSR-----   | 2132                      |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1876 | fydsd        | wnpmd               | aqadqchr               | igtrdvhlyl                              | iserteenil             | kanqkmlgdmaiegg           | 1935 |
| Qy | 2133 | -----        | PGKSP               | ERSHVSSEPE             | YPIPPQVPV                               | VHEKODSLLILLSORGAE     | 2174                      |      |
|    |      | :::          |                     |                        |                                         |                        |                           |      |
| Db | 1936 | nftay        | kqqtirel            | fdmple                 | psssvpsapee-----                        | eeetvas                | qthlil-----egal           | 1987 |
| Qy | 2175 | PABQR        | NDA                 | RSPGISY                | SLPFFTKLENTSP                           | MVSKKQEIF-----         | RKLNSGGGSDMAAA            | 2229 |

|        |          |                                                             |                      |
|--------|----------|-------------------------------------------------------------|----------------------|
| Db     | 1988     | raedeediraatq-----akaeqvaelaefnendgfpagegeaeagrpgaedeemstra | 2039                 |
| Qy     | 2230     | QPGTEIFNLPAVTTSGSVSSRGHSHFADPASNLGLEDIIIRKALMGSPD-----DKVE  | 2280                 |
| Db     | 2040     | e--qelaalveqlt--pieryamkfilea---leevsreelkqaeqveaarkldldqak | 2091                 |
| Qy     | 2281     | DHGVMVSMQMGVVPQTAN-TSVVTTGETRRREGDPSPHSGGVCKPKLISKNSRKSXSPI | 2339                 |
| Db     | 2092     | eevfrlpqeeegpgagdesscgtgggthr-----rskkakap--                | 2129                 |
| Qy     | 2340     | PCQGYLGTERRPSVSG-----VHSEGDYHRQTPG-WAWEDRPSPSTGSTQPPYNPLTWR | 2391                 |
| Db     | 2130     | -----epgtrvrseirlgaraetqganh--tpvisahqtrsttptprcpsarevpr    | 2179                 |
| Qy     | 2392     | MLSSPTPTPIACAPSAVNOQAAP                                     | 2413                 |
| Db     | 2180     | paprprrtp-asapaipalp                                        | 2200                 |
| RESULT | 31       |                                                             |                      |
| AA     | AA23963  |                                                             |                      |
| ID     | AA23963  | standard; Protein; 2783 AA.                                 |                      |
| XX     | AA23963; |                                                             |                      |
| AC     | AA23963; |                                                             |                      |
| DF     | DF       | 06-JAN-1993 (first entry)                                   |                      |
| DE     | DE       | AFP-1 (Ala 2460 Val).                                       |                      |
| XX     | XX       | Enhancer; alpha fetoprotein; homeodomain; zinc-finger.      |                      |
| KW     | KW       | Homo sapiens.                                               |                      |
| OS     | OS       |                                                             |                      |
| XX     | XX       |                                                             |                      |
| FH     | FH       | Key                                                         | Location/Qualifiers  |
| FT     | FT       | Domain                                                      | 1231..1291           |
| FT     | FT       | Domain                                                      | /note= "homeodomain" |
| FT     | FT       | Domain                                                      | 1328..1388           |
| FT     | FT       | Domain                                                      | /label= homeodomain  |
| FT     | FT       | Domain                                                      | 1727..1786           |
| FT     | FT       | Domain                                                      | /label= homeodomain  |
| FT     | FT       | Domain                                                      | 2032..2091           |
| FT     | FT       | Domain                                                      | /label= homeodomain  |
| FT     | FT       | Domain                                                      | 67..99               |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 123..155             |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 171..203             |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 306..366             |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 443..476             |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 484..543             |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 628..660             |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 679..711             |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 1066..1097           |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 1411..1442           |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 1613..1643           |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 1794..1825           |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 2107..2139           |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 2540..2571           |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |
| FT     | FT       | Domain                                                      | 2606..2638           |
| FT     | FT       | Domain                                                      | /label= zinc-finger  |







|           |                                                        |                                                              |    |      |
|-----------|--------------------------------------------------------|--------------------------------------------------------------|----|------|
| Qy        | 1889                                                   | SGKPPQHSSVVYSEACKDKGPPPKSRYE-----EELTRGKT-----               | TI | 1927 |
| Db        | 1608                                                   | fggprpr-----rqpppprpfyldsleekrkqrserierifqlseahgalapv        |    | 1658 |
| Qy        | 1928                                                   | TAAAFID-VIIITRIASDKDAREG-----SQSSDSSSLSSHRYETPSPAIE---       |    | 1974 |
| Db        | 1659                                                   | ygtevidctfpqvasipgrspshptfwtyeeahravlfpqqlgdlseierfi         |    | 1718 |
| Qy        | 1975                                                   | -VISPASSAP-----POEKIQTQVPEVVKANQAEENDPTQRYEGPLHHYRPQQESPSQ   |    | 2027 |
| Db        | 1719                                                   | fvmpveapppsIhachpppwlaprqaaafqeqlaselwprar---plhrivcmrmtgfpd |    | 1775 |
| Qy        | 2028                                                   | QQIAPP-----SSQAEGMGQVPRTHRLI-----TLADHTCQIIITQD---FA         |    | 2064 |
| Db        | 1776                                                   | lrliqydcgklqtlavllrqikaeg-----hrvliftqmrmldvleqfityghlyl     |    | 1828 |
| Qy        | 2065                                                   | RNOVSQTPQOPTSTFQNSPALYSTPVRNKT-----SNRYSPESQAO               |    | 2108 |
| Db        | 1829                                                   | rldgstrveqgalmernadkrifcflstrsgvgvnlgtadvvfysdwnptmdaq       |    | 1888 |
| Qy        | 2109                                                   | S-----VH-QRPGGRVSPENLVKSRGSR-----                            |    | 2132 |
| Db        | 1889                                                   | aqdrchriqtdrvhlyrlisertveenilkkanrkmlgdmaiegnfttayfkqqtir    |    | 1948 |
| Qy        | 2133                                                   | -----PKGPERSHVSSEPIEPISPQPVVHVEKODSLLLSORGAEPAEQRNDARSPGS    |    | 2187 |
| Db        | 1949                                                   | elfdmpleepsssvsapee-----eeetvaskthil---equalcraedeediraatq   |    | 2000 |
| Qy        | 2188                                                   | ISYLPSEFFTKLENTSPWVKSKQEIF-----RKLNSGGGSDMAAQPGTEIFNLPAVT    |    | 2242 |
| Db        | 2001                                                   | -----akaeqvaelaefnendfpagegeeaagrpgaedeemsrae--qeiaalveql    |    | 2050 |
| Qy        | 2243                                                   | TSGSVSRGHSFADPASNLGLEDIIRKALMGSEF-----DKVEDHGVMVMSQPMGVV     |    | 2293 |
| Db        | 2051                                                   | t-pieryamkfileas----leevsreelkaeeqveaarkdlqakeevfrfpqeeeg    |    | 2104 |
| Qy        | 2294                                                   | PGTAN-TSVVVTSGETRREEDPSPHSGGVCKPRLISKNSRKSKSPIPGOGYLGTERPSS  |    | 2352 |
| Db        | 2105                                                   | pgagdesscgtgggthr-----rskkakap-----erpgt                     |    | 2134 |
| Qy        | 2353                                                   | VSS-----VISEGDYHQRTPG-WAWEEDRPSSTGCTGFPPNPIITMRMLSTPPTPIACA  |    | 2404 |
| Db        | 2135                                                   | rvserlrgaraetgganh--tpvisahgrsttttpprcsparervpaprpptp-asa    |    | 2191 |
| Qy        | 2405                                                   | PSAVNQAAP 2413                                               |    |      |
| Db        | 2192                                                   | paalpavp 2200                                                |    |      |
| RESULT    | 33                                                     |                                                              |    |      |
| AAAR23962 |                                                        |                                                              |    |      |
| ID        | AAAR23962                                              | standard; Protein; 2783 AA.                                  |    |      |
| XX        | AAAR23962;                                             |                                                              |    |      |
| AC        | AAAR23962;                                             |                                                              |    |      |
| XX        | AAAR23962;                                             |                                                              |    |      |
| DT        | 06-JAN-1993                                            | (first entry)                                                |    |      |
| DE        | AFP-1.                                                 |                                                              |    |      |
| XX        | Enhancer; alpha fetoprotein; homeodomain; zinc-finger. |                                                              |    |      |
| KW        | Homo sapiens.                                          |                                                              |    |      |
| OS        |                                                        |                                                              |    |      |
| XX        |                                                        |                                                              |    |      |
| PH        | Key                                                    | Location/Qualifiers                                          |    |      |
| FT        | Domain                                                 | 1231..1291                                                   |    |      |
| FT        | Domain                                                 | /note= "homeodomain"                                         |    |      |
| FT        | Domain                                                 | 1328..1388                                                   |    |      |
| FT        | Domain                                                 | /label= homeodomain                                          |    |      |
| FT        | Domain                                                 | 1727..1786                                                   |    |      |
| FT        | Domain                                                 | /label= homeodomain                                          |    |      |
| FT        | Domain                                                 | 2032..2091                                                   |    |      |
| FT        | Domain                                                 | /label= homeodomain                                          |    |      |

|    |             |                                                                       |                                         |
|----|-------------|-----------------------------------------------------------------------|-----------------------------------------|
| Ft | Domain      | 67..99                                                                | /label= zinc-finger                     |
| Ft | FT          | 123..155                                                              | /label= zinc-finger                     |
| Ft | Domain      | 171..203                                                              | /label= zinc-finger                     |
| Ft | FT          | 306..366                                                              | /label= zinc-finger                     |
| Ft | Domain      | 443..476                                                              | /label= zinc-finger                     |
| Ft | FT          | 484..543                                                              | /label= zinc-finger                     |
| Ft | Domain      | 628..660                                                              | /label= zinc-finger                     |
| Ft | FT          | 679..711                                                              | /label= zinc-finger                     |
| Ft | Domain      | 1066..1097                                                            | /label= zinc-finger                     |
| Ft | FT          | 1411..1442                                                            | /label= zinc-finger                     |
| Ft | Domain      | 1613..1643                                                            | /label= zinc-finger                     |
| Ft | FT          | 1794..1825                                                            | /label= zinc-finger                     |
| Ft | Domain      | 2107..2139                                                            | /label= zinc-finger                     |
| Ft | FT          | 2540..2571                                                            | /label= zinc-finger                     |
| Ft | Domain      | 2606..2638                                                            | /label= zinc-finger                     |
| Ft | FT          |                                                                       |                                         |
| Pn | EP487229-A. |                                                                       |                                         |
| Xx | XX          |                                                                       |                                         |
| Xx | PD          | 27-MAY-1992.                                                          |                                         |
| Xx | PF          | 07-NOV-1991;                                                          | 91EP-0310334.                           |
| Xx | PR          | 07-NOV-1990;                                                          | 90JP-0301412.                           |
| Xx | PA          | (SNOW ) SNOW BRAND MILK PROD CO LTD.                                  |                                         |
| Xx | PI          | Higashio K, Morinaga T, Tamaoki T, Yasuda N;                          |                                         |
| Xx | DR          | WPI; 1992-176828/22.                                                  |                                         |
| Xx | DR          | N-PSDB; AAQ24828.                                                     |                                         |
| Xx | PT          | DNA encoding protein binding to alpha-fetoprotein gene enhancer -     |                                         |
| Xx | PT          | useful for prodn. of biological active protein                        |                                         |
| Xx | PS          | Claim 1; Page 7; 24pp; English.                                       |                                         |
| Xx | CC          | This protein specifically binds to the enhancer of the alpha-         |                                         |
| Xx | CC          | fetoprotein gene. The DNA encoding this protein may be useful for     |                                         |
| Xx | CC          | the prodn. of biologically active proteins by inserting the DNA into  |                                         |
| Xx | CC          | an expression vector and co-transfecting animal cells with another    |                                         |
| Xx | CC          | expression vector in which a gene for the biologically active protein |                                         |
| Xx | CC          | is placed under the control of alpha-fetoprotein gene enhancer and    |                                         |
| Xx | CC          | promoter.                                                             |                                         |
| Xx | SQ          | Sequence                                                              | 2783 AA;                                |
|    |             | Query Match                                                           | 2.6%; Score 326; DB 13; Length 2783;    |
|    |             | Best Local Similarity                                                 | 17.8%; Pred. No. 3.6e-09;               |
|    |             | Matches 514; Conservative                                             | 343; Mismatches 1014; Indels 1016; Gaps |
| Qy | 39          | AVPDRSRSHLEVSQAOLQQOQQQLRRR--PSLISE-----FHFGSDRPOE---84               |                                         |
|    |             | : : : : : : : : : :                                                   |                                         |
| Db | 391         | avpd-rdgnsnleagk--qpetstedlgknlpstategsgdlkpdpdgvsredsgfi 447         |                                         |
|    |             | : : : : : : : : : :                                                   |                                         |
| Qy | 85          | --RRTSYEPFHPGSPVDH--DSLESKRPRLEQVSDSH-----118                         |                                         |
|    |             | : : : : : : : : : :                                                   |                                         |
| Db | 448         | cwkkgcngvktksaalqthfnevhakrpgl-pvsdrhvkvrcngcgslafktiekqlbs 506       |                                         |

```

Query Match 2.6%; Score 326; DB 13; Length 2783;
Best Local Similarity 17.8%; Pred. No. 3.6e-09;
Matches 514; Conservative 343; Mismatches 1014; Indels 1016; Gaps 129;

Qy 39 AVPYRSHLEVSQASQLLQQQQQLRRR--PSLLSE-----FHPGSDRPOE--- 84
||||| :
Db 391 avpd-rdgnsnleagk-qpetsedlgnlpaaasteqsgdlkpsadpgsvredsgfi 447
||||| :

Qy 85 --RTSTSEPPFPGSPVDH--DSLESKRPRLRQVSDSH----- 118
: :
Db 448 cwkkgcnqvftsaalqthfnvhaikraql-pvsgrhvvkvrncacslafkieklqghs 506

```





QY 482 KKNENYA-----LVRNNGK-----RRGRNOQIARPSOEKVEE-----KEEDKAETE 526  
Db 168 kdkneksdvtql1tvkdegagaaavagdhqe-----psvetavgesakeselkyste 224  
QY 527 KKEEKKDEEKEDE-----KEDSKENTKEDKIDGTAETETEEQATPRGRKTANSQGRRMG 583  
Db 225 kqegtlkqedssteip1qaesdgaeeekdeg--eekqeke----- 264  
QY 584 RITRSMYNAAAASAAAAATEEPPLPPPPPISTETPEVETSWTBEEMEVAKGLVHEH 643  
Db 265 -----ptksp-----espspvnsettss-----kffth 290  
QY 644 GRNWAATAKMGVTGKSEACKNFYNYKRRHNLNLLQOHKOKT-----SRKPREERDVQ 698  
Db 291 g--wagwrkktfskksdledlletaeakrkeqaeakvdeeketepaseeqeapadtqar 348  
QY 699 CESVASTVSAQ--EDE--DIEASNEE-----NPDESEVA-----VKPSEDSPEN 740  
Db 349 lsadyekvelpledqvgdleaaseekcaplatevfdekmeahqevvaevhstvtekteee 408  
QY 741 ATSRGNTPEPAVELEPTTETAPSTSPSLAVSTKPAEDESVEVETQVNDISABTAQOMDVQ 800  
Db 409 qggggaeggvvvegtgeslp--peklaepqevpgeapaeelmk-----sremcvsg 459  
QY 801 QEHAESGVCDDPPATKADSDVDVEVRVPHENHASKVBDNTER-----D 845  
Db 460 gdht-----qitdlspek-----tlphpegivsevmsslerikvqgspklkfssg 510  
QY 846 LDRASEKVEP-----RDELDVVAQIQAORPEQSDNDSSATCSADE----- 888  
Db 511 lkk1sgkkqkgkrgggdeegpgeyghintesadesqkessasspeeeettclekpg 570  
QY 889 -----VDEPERQRMFPMDKPSLLNPTGSLVSPKPNLDLPQLQHRA 934  
Db 571 leapdgaeeegttsdgekregitpwasfkkmvtpkkrv-----rrpsed-----ke 619  
QY 935 AVIPMVSCPTCNPIGTPVSGYALYORHIKAMHESALLEORQORQOIQIDECRSSSPC 994  
Db 620 eelekvksat-----lastdsvemqdevktvgeekpkepkrrvd----- 661  
QY 995 GTSKSPNREWEVLPAQHILITNLPGLVRLPTTRTPPPPLIPSSKTTVASEKPSPTMG 1054  
Db 662 -tsvs-----weal-----icvgskkrrarkasssddeg 689  
QY 1055 GSISQGTPTCYLTSH--NQASYTOETPKPSVGSISLGLPQOESAKSATLPYIKQEFSP 1112  
Db 690 gprtlgg-----dshraeesakkeagtdavpast-----qeqdaqgsssp-----ep 733  
QY 1113 RSONSQPEGLIVRAQHEGVV---RGTAQIQEGSITRGTPTSKISVESIPSLRGSITQGT 1169  
Db 734 aglsegegstwesfirlvtprkkskieeka--edssveqlsteiepsreeswvsik 791  
QY 1170 PALPQTGIPTEALVKGISIRMPIEDSPKREPAASKGHVYEGKSGHILSYDNK--- 1226  
Db 792 kfipgr---rkradgkqeatvedsgpvneiddpnpvavv-----plseynaverek 842  
QY 1227 -NAREGTFSPR-----TAHEISL-----KRSYSEVEGNIKOGMSRESFVS 1266  
Db 843 meadgnntelpqllgavvyvseelskltvhtsvavldtravtsve-----erspswisavt 899  
QY 1267 APLBGLCRALPRGSPHDLKERNVLGSGIMQGTTPRATTESFEDGLKYPKQIKRESPIIR 1326  
Db 900 eplehtageamp---pveevtekdi-----laetpvtlqt1pegkdahdmvtse---vd 949  
QY 1327 AFEQAITYKGPYDGIITIKEMGRSIIHEIPQDILITQBSRKT--PEVQSTRPIIEGISIQ 1384  
Db 950 fteavtatetseairt--eevteasgeettmdmsavsgitdspttteeatpve--ves 1006  
QY 1385 GTPIKFNNSQSAIKHNKSLITGSKL-----SRGMPPLEIVPENIKVVERGK 1434  
Db 1007 gvidteeeertqallqavadkveesqvpatqvtqvtgtskaiekveeedsevlasek 1066

QY 1435 YEDVKAGETVRSRHTSVVSSG-----PSV-----LRSTLHEAP 1467  
Db 1067 ekdvmpkqpvqeaagahlaqsgsetgqatpeslevpervtadvhvatcqviklqlmeqav 1126  
QY 1468 KAQLSPGIYDDTSARRTPVSYQNTMSRGSPMMNRTSDVTI--PPNKTSTNHERKSTLPTQ 1525  
Db 1127 apsesetltdsetngstpladsdtad-----gtqgdetidsqskataavrgsqv--te 1178  
QY 1526 RESIPAKSPVGVDPVVSHPFDPHHRGSGTAGEVYWSHLPTQLDPAMPFHRLDPAALV 1585  
Db 1179 eeaataakeepstlp--nnvpaqeeh--geepgr-----dvleptqgeltaaavpv 1225  
QY 1586 LFORQLSPTPCYPSQYQLYAMENTRQILNDYITISQOMQNLNRPDVARGLSRQPLGLP 1645  
Db 1226 laktev---gqgeev-----dwldgekvk-----eeqevf-- 1253  
QY 1646 YPATRGIDITNMPPTILVPHPGGTSTPPMDRITYIPCTQITTFPPRPNYSASMSGPHPTH 1705  
Db 1254 -----hsgpnsqkaad-----vtvdsevmgvacqcekeste 1284  
QY 1706 LAAAASAEERERERERERERERERERERERERERERERERERERERERERERERER 1765  
Db 1285 vqslsleegemetdvekeketk-----peqvseege-----qe 1318  
QY 1766 TMLQORPSVFGTNGTSVITPLDPTAQLRIMPLPAGGSPISQGLPASRYNTAADALAAV 1825  
Db 1319 taapeh-----egtygkpvlt--ld-----mpssergkalgsig-- 1350  
QY 1826 DAAASAPQMDVSKTKESKHEAARLEENLRSAASAVSQOOLEQKTLVEKRKSVQCLYTSS 1885  
Db 1351 -gspslpdqg-----kagclevqvgsltdttvtqtaeavekie----- 1387  
QY 1886 AFPSGKQPQSHSVVYSEAGDKG-----PPPKSRVEEELTRGKTTI----- 1927  
Db 1388 -----tvvisetgespecvghllpaekss-----atgghwtlghaedvtplgp 1431  
QY 1928 -TAANFIDVIT-----RQIASDKDARERGSQSDSSSSLSHRYE 1967  
Db 1432 esqaespiivtpapestlhpdlqgeisaqrseeeedkpadgpadgkestaiekvlk 1491  
QY 1968 TPSDAIEVSP-----ASSPAPPOEKLQTYQYQPEVVRANONENDP-----T 2007  
Db 1492 aepeillelesksnivinvigtavdqfartetapethaydsqtqvpacridsrepcrwt 1551  
QY 2008 RQYEGPLHHYRQOESPSPQOQLPPSSQAEQMGVPR 2044  
Db 1552 kmkdakmhvpq-----predlqvlvleawaq-pr 1582  
  
RESULT 35  
AAW53863  
ID AAW53863 standard; peptide; 1780 AA.  
XX  
AC AAW53863;  
XX  
DT 13-JUL-1998 (first entry)  
XX  
DE Human gravin polypeptide.  
XX  
KW Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;  
KW cAMP-dependent protein kinase; protein kinase C; autoimmune disease;  
KW Myasthenia gravis; nicotinic acetylcholine receptor.  
XX  
OS Homo sapiens.  
XX  
PN US5741890-A.  
XX  
PD 21-APR-1998.  
XX  
PF 19-DEC-1996; 96US-0769309.  
XX  
PR 19-DEC-1996; 96US-0769309.  
XX



(UYOR-) UNIV OREGON HEALTH SCI.

PA Klauck TM, Nauert JB, Scott JD;

PI WPI; 1998-260552/23.

DR N-PSDB; AAV23545.

XX

PT New polypeptide fragments of protein kinase binding protein gravin -  
PT are useful for the study of modulation of action between gravin and  
PT protein kinase(s)

XX

PS Example 1; Column 19-32; 32pp; English.

XX

CC This sequence corresponds to the human gravin polypeptide, and represents  
CC a polypeptide of the invention. The polypeptides are fragments capable of  
CC binding to type II regulatory subunit of cAMP-dependent protein kinase  
CC (PKA). Gravin is a kinase anchoring protein that binds to type II  
CC regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an  
CC antigen of the autoimmune disease Myasthenia gravis (MG), where a patient  
CC develops antibodies against their own nicotinic acetylcholine receptors.  
CC The polypeptides are useful for providing analogues of gravin in the  
CC study of the modulation (e.g. blocking, inhibiting and stimulating) of  
CC interactions between gravin and kinase. The peptides are involved in the  
CC modulation of gravin-kinase interactions.

XX

SQ Sequence 1780 AA;

Query Match 2.68; Score 323.5; DB 19; Length 1780;

Best Local Similarity 18.68; Pred. No. 2.7e-09;

Matches 361; Conservative 273; Mismatches 741; Indels 565; Gaps 82;

QY 496 KRRGRNQIARPSQEEKV-----BEKEDKAETKEEKKDEEKDEKEDSKENT 548

DB 83 GQKGAALGQALNSQEEVIVTEVGRQDSVSDKEMATKSAVVHIDTQGEEN 142

QY 549 KEKIDGTAETTEREQAPRGKRTANSQGRK-----GRITR 587

DB 143 RNLQIPSESNLEELTQPT---ESQANDGFKVFKVGFVFKVKKDKTEKPTVQLIT 199

QY 588 SMTNEAAAAASAA-----AAAATEPPPPPLPPPPPEISTE-----PVETSR 627

DB 200 VKDEGEAGAGAGHQDPSLGAGAAKSEKPEKSTKEPKETILKRESHAELSPAESGQ 259

QY 628 WTEP-EMEVAKGLVHERNMAAIKAVGKSEACQKNFYFNRRRLNLLQHQHOKT 686

DB 260 AVECKEKEGKEKESKSAESPTSPVTSSTGTFKFF-----TGWAGWRKTS 311

QY 687 SRKPREEDVSQCSVASTVSAQDEDEIASNEENPEDSEVAVKPSEDSPENATSRGN 746

DB 312 FRKPKK-----DEVEASEKKKEQEKVDTEEDGKAEVASEKLTASEGHPQEPAES 364

QY 747 TEP-----AVELEPTTETAPSPSLAVPSTKPAE--DESEV----- 781

DB 365 HEPLISAQYKVELPSEQVSGSGGSPSEKPAPIATEVFEKIEVHQEVVAEVHSTVE 424

QY 782 -----TQVND-----SISATAEQMDVQQRH-----SAEAGVCDPPPPATKAD-SVD 823

DB 425 ERTEEQKTEVEETAGSPAAELVGMDEAPQEAEPKELVKIKETCVSGEDTQGDADLSPD 484

QY 824 VEV--RVPENHASKVEGDNTER-----DLDRASEKVEP-----RDEDLV 861

DB 485 EKVLSPKEGVVSEVEMLSQGRMKVGSPLKIFTSTGLKKLGGKKGKGRGGGDEESG 544

QY 862 VAQINQARPEPQSDNDSSACADE-----VDEGERQ 896

DB 545 EHTGVADSPDSQGEQKESASSPEELTCLKGLAEVQDGEAEAGATSQDEKKRE 604

QY 897 RMFPMWCKPSLLNTGTSILVSSPLKPNPLDLPLQLQRAAVIPPMVWCTPCNIPITGVPSG 956

DB 605 GVTWASFKKMVT----- 618

QY 957 YALYQRHIKAMHESALLEEQRQEQIDLECRSSTSPCGTSKSPNREWEVLQAPAPHLIT 1016

DB 619 ---KRVIRPSES-----dkedeldkvsatlsstestaseemqemkgsveepkee--- 667

QY 1017 NLPEGVRLPTRTRPPPPPLIPSSKTTIVASEKPSFIMGSGISQGTPTGTYLLSHNQASVQ 1076

DB 668 -----pkrvdtsvswaelicvgsakk-----farrirssde 699

QY 1077 ETRKPSVSGISLGLPROQESAKSATLPYIKOEEFSPRSONSQPEGLLVRQAHEGVVRCGA 1136

DB 700 eggpkamgg-----dhqkade-agkdktgtdgilagsqehpgqgss 741

QY 1137 GAIQEGSITRPTPSKISVESIPSLRGSITOGTTPALQGTGPTALVKGSISRMPIEDSS 1196

DB 742 spedags-----ptegegvswesfkrivtrkkskskleeksedsiaags-----gvehst 792

QY 1197 P--EKGREAAKSHVIEGSGHLSVDNKNKAREGTRSPRTAHEISLKRSYSEVEGNI 1254

DB 793 pdepgkee-----swvsikkfipgrkkrrp-----dgkqeqapvedag 831

QY 1255 KQGMRESPPVSA--PL-----EGLICRALPRGSPHSDLKERTVLGSGIMQGTPTAT 1304

DB 832 ptganeddsdvpavvpIseydaverkmeaqagkaeqpeqkaatevskelseeqvhhmm 891

QY 1305 TESFEDGLKTPKQIKRESP-----PIRAP--EGAITKGKPYDGIITTKEMGRSITHE 1353

DB 892 aaavadgttraatiieerspswisasvtepleqveaeaaallteevlereviaeeepptvte 951

QY 1354 -IP-----RODILTQESRKTPTEVQSTRPIIEGSIQGTPIKFDNNSGQSAIKHN----- 1402

DB 952 pipeneargdtvvsaealtpeavtaa-----etaggIpgseegteasaeetems 1003

QY 1403 -VKSLLITGPSKLSGMPLEI---VPENIKVVERKGYEDVKA-GETVRSRHTSVSVSSGP- 1456

DB 1004 avsqldspdtteatpvqevegvpd-iesqerrtqevlqavaekveesqlptgtgpe 1062

QY 1457 SVLRSTLH---EAPKAQL-SPGIYDDT-----SARRTPSVYQNTMSRSGSP-MMNRT 1502

DB 1063 dvlqpvdraeaeerpeeqaeasglkktadvlkdadeakteptfgkvvgvgttpefesa 1122

QY 1503 SDVT--TPNNKSTNHERKSTLTPTQRESIPAKSPVPGVDVPSVSHSPDPHRRHGTAGEVY 1560

DB 1123 pqvteslessevlvtcqaetlagvksqemveqaip---pdsvetpdtsetdgtspvadf 1179

QY 1561 WSHLPTQLDPAMPFHRA---LDPA---AAAYLFQRLSPTPGYPQYQIYAMENTRQT 1612

DB 1180 dapgttkdeiveiheenevhlvpvrgteaevpaqkerppa---psaf-vfgeetkeqs 1235

QY 1613 ILNDYITISQOQVNLRPDVARGLSPR-----EQPLGLP-YPATRGIIIDLTNMPP 1660

DB 1236 kmedtlehtdkevasv--etvsilsktegtqeadgvdktkdvffeglegsld-tgi-- 1290

QY 1661 TILVPHPGGTSTPPMDRITYI---PGTOITFPPR---PYNASMSPGHPHTLAAAASA 1712

DB 1291 -----tvrekvtvalkgteaeackddalelqshakspsspveremvqv 1339

QY 1713 ERER-ERE---REKERERERIAAASDLYLRPGSEQ-----PGRPGSHGYVRSP 1757

DB 1340 erkteaeapthvneekiehetavtveev-----skllqtvnvpiidgakevslegsp 1394

QY 1758 SPVSVRTQETM-----LQORPSVFGQGTNGTSVITPLDPTAQIRIMPL-----PAGGPSISQ 1807

DB 1395 ppcIqgeeeavctkiqvqssseasfItaaeeekvlgeta--niletgetlepagahlvle 1452

QY 1808 GLPASRYNTAA---DALAALVDAASAAPQMDVS-----KTKESKHEAAR 1848

DB 1453 eksseknedfaahpgedavptgpcqakstpvivsattkklgslsdlegektslkwsde 1512

QY 1849 LEENLRSRAVS---EQOQLEQKTLVEKRS-----VQCLYTSASAPSG 1890

DB 1513 vdeqvacekvsvaiedlepengileletksklvniqltavdqfvrteetemtls 1572

QY 1891 KPQPHSSVYVSEAGKDGPPPKSRYEE-ELTRGKTTTITTAANFIDVITRITIASDKDARE 1949

Db 1573 elqtqahvikads-qdagetekegeepqasagdetpitsa----- 1612

Qy 1950 RGSQSSSSSLSHRVTETPSDAIEVTSIPASSPAPPQEKLTQYQPEVVKANQANDPTRQ 2009

Db 1613 --keesestavqgah-----sdlakdmseas-----ekmtveve-----gstvnd--qg 1653

Qy 2010 YEGPLHRYRPOQSPSPQOOLPPSSQAEGMG--QVPRTHRLITLADHICQIITO----- 2061

Db 1654 le-----evlpseeeeggagtksvpeddghallaerleksiivepkekek 1698

Qy 2062 ----DFARNQVSS-----QTPQOQPTSTFQNSPS-----ALVSTPVRTKTS 2098

Db 1699 gddvdpnqnsaladitdasggltkespd-----ngpkqekedagevelqegkvhsead 1754

Qy 2099 NRYSPESQAQSVHHQRPGSR 2118

Db 1755 kaftpqaeelqkqeresak 1774

RESULT 36

AAB15380

ID AAB15380 standard; Protein; 1780 AA.

XX AC AAB15380;

DT 26-JAN-2001 (first entry)

XX DE Human gravin protein sequence.

XX KW Human; gravin; PKA RII binding site; myasthenia gravis;

XX KW kinase anchoring protein; CAMP dependent protein kinase.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Binding-site 265..556

FT /note= "PKC binding site"

FT Binding-site 1526..1582

FT /note= "PKA RII binding site"

FT Region 1537..1563

FT /note= "PKA anchoring site"

XX US6090929-A.

PN 18-JUL-2000.

XX PD 19-DEC-1997; 97US-0994570.

XX PF 19-DEC-1996; 96US-0769309.

XX PS (UYOR-) UNIV OREGON HEALTH SCI.

XX PA Klauck TM, Scott JD, Nauert JB;

XX PI WPI; 2000-523763/47.

XX DR N-PSDB; AAA74903.

XX Novel polynucleotides useful for detecting gravin in patients suffering

PT from Myasthenia gravis encodes CAMP-dependent protein kinase-binding

PT polypeptide and protein kinase C-binding polypeptide of gravin

XX Claim 1; Column 35-45; 34pp; English.

XX The present sequence is the protein sequence of human gravin. Gravin is

CC an A-kinase anchoring protein (AKAP) which is involved in the

CC localisation of CAMP dependent protein kinase A (PKA) via interactions

CC between the RII binding region and the PKA regulatory subunit RII. Gravin

CC is also an antigen found in myasthenia gravis sufferers, and it is

CC thought that antibodies to it may be useful in modulating the binding of

CC PKA, and thus aid in the treatment of the disease. The gravin coding

CC sequence was isolated by first screening a human umbilical vein

CC endothelial cell cDNA library with serum from a myasthenia gravis

CC patient, and then searching a human heart cDNA library for sequences

CC resembling the isolated sequence. This was done because the first

CC sequence obtained was shown to be shorter than the full length cDNA.

XX SQ Sequence 1780 AA;

Query Match 2.6%; Score 323.5; DB 21; Length 1780;

Best Local Similarity 18.6%; Pred. No. 2.7e-09;

Matches 361; Conservative 273; Mismatches 741; Indels 565; Gaps 82;

Qy 496 GKRRGRNQIARPSQEEKV-----EKEEDKAETKEKEEKKDEEKKDEKESKENT 548

Db 83 gqgqalngqalnsqeveevivtevgqrdsevdvserdskematksavvhditdgqeen 142

Qy 549 KEKDIDGTAEETEERQATPRGKKTANSQGRK-----GRITR 587

Db 143 rnieqibssesnleeltqpt---esqandigfkvfkfvgfkftvkdkdtekptdvqlt 199

Qy 588 SMTNEAAAAASAA-----AAATEPPPPPLPPPPPISTE-----PVETSR 627

Db 200 vkdegegaagdhqpslgagaasasepkstekteetlkregshaieispaesgq 259

Qy 628 WTEP-EMEVAKGLVHERNMAAIAMVGTAKSEACKNFYFNKRRHNLNLLQOHKQKT 686

Db 260 aveckegeekqekpskaesptsvtsetgstfkff-----tgwagwrkts 311

Qy 687 SRKPRERDVSQESVASTVSAQDEDEIEASNEEENEDSEVEAVKPSDESPPENATSRGN 746

Db 312 frpkpe-----deveasekkkeqekvdeedgkaevaseklitaseahqppeasa 364

Qy 747 TEP-----AVELEPTTETAPSTPSLAVPSTKPAE--DESV----- 781

Db 365 heprlsaeeyekvelpseeqvgsgpseeekpaplplatevfekievhqeevaevhvstve 424

Qy 782 -----TQVND-----SISATAEQMDVQDEH-----SAEGSVCDPPPATKAD--SVD 823

Db 425 erteekteveetagsvpaeelvgmdaepqeaepakelvkiketcvsgedptggadispd 484

Qy 824 VEV--RVPENHASKVEGDNTKER-----DLDRASEKVEP-----RDEDLV 861

Db 485 ekvlskpppegvvsevmssqermkvqgspklkftstgklksgkqkgrgggdeesg 544

Qy 862 VAQINAQRPQSDNDSDSATCSADE-----VDGEPEPQ 896

Db 545 ehtqvpadspdsgeekgessasspeeeitclekglaevgqdgaeagatsdgekkre 604

Qy 897 RMFPMDKPSLLNPTGSLYSSPLKPNPLDLPLQOHRAAVIPPMVSVCTPCNIPGTGTPVG 956

Db 605 gvtpwaskkmvtp----- 618

Qy 957 YALYORHIKAMHESALLEEQROEQIDLECRSSTPCGTSKSPNREWEVLQAPAPHOLIT 1016

Db 619 ----kkrvirpses-----dkedeldkvksatlstestaseemqemkgsvveepkee--- 667

Qy 1017 NLPEGVRLPTRRPPPLIPSKTTTVASEKPSFINGGSIQCTPGTYLTSHNOASYTQ 1076

Db 668 -----pkrvdtsvswaelicvsgsk-----rarrissde 699

Qy 1077 ETPKPSVGSISLGLPROQESAKSATLYIKOEESPSQNSQPEGLLVRAQHEGVVRGTA 1136

Db 700 eggpkamgg-----dhqkade--agdkdetgtdgilagsqehdpqggss 741

Qy 1137 GAIQEGSITRGTPTSKTSVSEIPSLRGSITQGTAPALQGTIPALVKGSISRRPIEDSS 1196

Db 742 speqags-----ptegegvatwesfkrlyprkkskleeeksedsiags-----gvehst 792

Qy 1197 P--EKGREEAASKGHVIEYEGSGHILSYDNINKNAREGTRSPRTAHEITSLKRSYSEVSGNI 1254

Db 793 pdepepgee-----Swvsikkfipgrkkrip-----dgkqeqapvedag 831

Qy 1255 KQGSMSRSPVSA--PL-----EGLICRALPRGSPHSDLKERTVLGSGIMOGTTPRAT 1304

Db 832 ptganeddsdvavvpviseydavekmeaqgagqaeqpedkaatevskseisqvhmm 891

|    |      |                                                                  |      |
|----|------|------------------------------------------------------------------|------|
| Qy | 1305 | TESFEDGLKVPKQIKRESP-----PIRAF--EGAITGKPYDGTITIKEMGRSIHE          | 1353 |
| Db | 892  | aaavadtgtraatifeerspswisasvtepleqveaeaaallteevleriveiaeeepvtve   | 951  |
| Qy | 1354 | -IP-----RQDILTOBSRKTPEWOSTRPIIBGSIQGTPIKFDNNSSQSAIKHN-----       | 1402 |
| Db | 952  | plpenreargdtvvsaeaeltpaavtaa-----etagplgseegteaasaeeetmvs        | 1003 |
| Qy | 1403 | -VKSILITGSPKJUSRGMPPILEI---VPENIKVVVERGKYEDVKA-GETVRSRHTSVVSSGP- | 1456 |
| Db | 1004 | avsqlltdpdtteeatpvcveggvpyd-leeqerrttgevlqavaekvkeesqpgtggge     | 1062 |
| Qy | 1457 | SVLRSYLH---EAPKAQL-SPGIYDDT-----SARRTPVSQYNTMSRGSP-MMNRT         | 1502 |
| Db | 1063 | dviqpvgraeerpeeqaeasglkktedvilkvdaeqaktepftggkvvgqgttpefeka      | 1122 |
| Qy | 1503 | SDVT--IPPNKSTNHBRKKTTLPTQRESIPAKSPVPGVDPVVSHSPDPHHRGSTAGEVY      | 1560 |
| Db | 1123 | pqytesieasselvttoqaetlagvksqemvmeqaiP---pdsvetpldsetdstpvdaf     | 1179 |
| Qy | 1561 | WSHLPTQLDPAmpFHRA-----LDPA---AAAYLFQRLSPTPGYPQOYLAYMENTROT       | 1612 |
| Db | 1180 | daggttqkdeiiveheenevhlvpvrgteaevpaqkerppa---psfv-fveetkeqs       | 1235 |
| Qy | 1613 | ILNDYITSQMQVNLRPDVARGLSRP-----BQPLGLP-YPATRGIIIDLTMNPP           | 1660 |
| Db | 1236 | kmedtihtdkdevs--etvsilsktegtqeadyadektdkdvffeglegsid-tgi--       | 1290 |
| Qy | 1661 | TILVPHPGGTSPPMDRIYI---PCQTITFPPR---PYNASMSPGHPHTHLAAAASA         | 1712 |
| Db | 1291 | -----tvrekvtvalkgtegeaeackdddalelqshakspvpervmvqv                | 1339 |
| Qy | 1713 | ERER-ERE---REKERERIAAASDLYLRPGSEQ-----PGRPGSHGYVRSP              | 1757 |
| Db | 1340 | erekteaeptnneeklehetavtvsvev-----skllqtvnvpilidgakevsslegsp      | 1394 |
| Qy | 1758 | SPSVRTQETM-----LQORPSVFOGTNGTSVITPLDPTAQLRIMPL-----PAGGPSISQ     | 1807 |
| Db | 1395 | ppclgqeeavctkiqvgseasftltaaeeekvlgeta--niletgetlepagahivle       | 1452 |
| Qy | 1808 | GLPASRYNTAA-----DALAALVDAASAAPQMDVS-----KTKESKHEAAR              | 1848 |
| Db | 1453 | ekseknedfaahpgedavptgpcqakstpvivsattkglssdlegekttslkwsde         | 1512 |
| Qy | 1849 | LEENLRSAAYS---EQOOLEOKTLEYEKRS-----VOCLYTSSAPFSG                 | 1890 |
| Db | 1513 | vdeqvacqevkvsvaldlepengilleetksklvqnliqtavdqvfvrteeatemlts       | 1572 |
| Qy | 1891 | KPQPHSVVYSEAGKDGPKPPKSRYEE-ELTRGKTTITAAFNIDVITRIQIASDKDARE       | 1949 |
| Db | 1573 | elqtqahvikads-qdaqetেকেপেপqasagadtetsa-----                      | 1612 |
| Qy | 1950 | RGSQSDSSSSLSHRYETPDSAEIVISPASPPAPQPKLOTYOPEVVVKANQAEENDPTRQ      | 2009 |
| Db | 1613 | --keesestavqgah-----sdiskdmseas-----ekmttveve---gstvnd--qg       | 1653 |
| Qy | 2010 | YEGPLHHYRPQESPSQOQLPPSSQAEGMG-QVPRTHRLITLADHICILITQ-----         | 2061 |
| Db | 1654 | le-----evvipseeeggagtksvpeddghallaelekslvpepkedek                | 1698 |
| Qy | 2062 | -----DFARNQVSS-----QTQPPQPTSTFNQSPS-----ALVSPVPRTKTS             | 2098 |
| Db | 1699 | gdvddpennqnsaladtdasggltkespdtt---ngpkqkekedaqaveileqekvhsesd    | 1754 |
| Qy | 2099 | NRYPSPESQAQSVHHQRPGSR                                            | 2118 |
| Db | 1755 | kaipqageelqkqeresak                                              | 1774 |



PT chromosome 11

Disclosure; Page 29-50; 90pp; English.

This sequence is encoded by the acute lymphoblastic leukemia (ALL-1) gene of chromosome 11. The ALL-1 gene was isolated by translocation breakpoint mapping. Fragments of the ALL-1 cDNA may be used to identify chromosomal abnormalities within the ALL-1 gene. These fragments may be used in the treatment and diagnosis of human leukemias such as acute lymphocytic, myelomonocytic, monocytic and myelogenous leukemia. ALL-1 protein shows three regions of homology to the Drosophila trithorax protein. These regions show 64%, 66% and 82% similarity respectively, to the Drosophila gene. The third region of homology constitutes the extreme C-terminus of the two proteins, both of which end in an identical sequence. The first homology region is cysteine-rich and contains sequence motifs analogous to four zinc finger domains (3-6) within the trithorax gene. The second region of homology is also cysteine-rich and corresponds to zinc fingers 7 and 8 of the Drosophila gene. The multiple conserved cysteines and histidines at the 3' end of the motifs allow two or three arrangements of the putative fingers. The structure of these cysteine-rich domains appears to be unique to the trithorax and ALL-1 genes.

| AA | Sequence | 3910 AA; |
|----|----------|----------|
| SQ |          |          |

Query Match 2.5%: Score 316: DB 14: Length 3910:

Query match 2.5%; score 310; DB  
Best Local Similarity 17.7%; Pred. No. 2e-08;

Matches 581; Conservative 367; Mismatches 1067; Indels 1264; Gaps 153;

Qy 13 FSTEQSRYPHSVQ--YTFPNTRHQQEFAPDYSRSHLEVSQASQLLQQQQQQQLR-RRP 69

```

366 fiededyppikarlestpnr---fapscsse-kssaaqhsqmsdsrscssp 420
Db

```

QY 70 SL-LSEFFHPSDRPQ---ERRTSYEPFFHGPSPVD---HDSLESKRPRLEQVSDSHFQRV 122

Db 421 svatstdsqaseelqvlpeersdtpvehp-plp1sqspenesndrrrrysvrsfgr 479

Qy 123 SAAVPLVHPLPEGLRASADAKKDPAFGGKHEAPSPISGQPCGDDQNASPSKLSKEELI 182

```
Db 480 ttxkxstlqsapq--getssppppll-----tpppl--qpas:sdhtpwm----- 524
```

|    |     |                |               |          |                  |          |     |
|----|-----|----------------|---------------|----------|------------------|----------|-----|
| QY | 103 | QSMKRVADREIARV | QQLALAKAAQQQL | EEEEAAAP | -----EPKAPVSPFFV | EQAKRSIV | 233 |
| D6 | 525 | -----          | -----         | -----    | -----            | -----    | 551 |

|     |               |       |
|-----|---------------|-------|
| 50  | -----ppci     | ----- |
| 51  | -----praspil  | ----- |
| 52  | -----pastap   | ----- |
| 53  | -----mqgkxslr | ----- |
| 54  | -----         | ----- |
| 55  | -----         | ----- |
| 56  | -----         | ----- |
| 57  | -----         | ----- |
| 58  | -----         | ----- |
| 59  | -----         | ----- |
| 60  | -----         | ----- |
| 61  | -----         | ----- |
| 62  | -----         | ----- |
| 63  | -----         | ----- |
| 64  | -----         | ----- |
| 65  | -----         | ----- |
| 66  | -----         | ----- |
| 67  | -----         | ----- |
| 68  | -----         | ----- |
| 69  | -----         | ----- |
| 70  | -----         | ----- |
| 71  | -----         | ----- |
| 72  | -----         | ----- |
| 73  | -----         | ----- |
| 74  | -----         | ----- |
| 75  | -----         | ----- |
| 76  | -----         | ----- |
| 77  | -----         | ----- |
| 78  | -----         | ----- |
| 79  | -----         | ----- |
| 80  | -----         | ----- |
| 81  | -----         | ----- |
| 82  | -----         | ----- |
| 83  | -----         | ----- |
| 84  | -----         | ----- |
| 85  | -----         | ----- |
| 86  | -----         | ----- |
| 87  | -----         | ----- |
| 88  | -----         | ----- |
| 89  | -----         | ----- |
| 90  | -----         | ----- |
| 91  | -----         | ----- |
| 92  | -----         | ----- |
| 93  | -----         | ----- |
| 94  | -----         | ----- |
| 95  | -----         | ----- |
| 96  | -----         | ----- |
| 97  | -----         | ----- |
| 98  | -----         | ----- |
| 99  | -----         | ----- |
| 100 | -----         | ----- |

Db 552 eptfrwtslkhrsepqyfsakyaqeglirkpfdnfrpp---pl--tpedvgfsgfs 606

QY 273 -----KYHENIKTNQVMRKLLI---LFFKRNHARKQREQKICQRYDQIMEAW 318

```
Db 607 asgtaasarlfsplhsgrtfmdmhrsllraprfpseahr-----ifesv 659
O:" 310 EFVVDPIENNDDBYAKKSCWMBBVBVVAVDFEETDVAADPACACTMTATGCHN 370
```

[illegible]

QY 379 EISEIIDGLS-----EQENNEKMRQLSVIPPMFDAE 411

Db 710 ppsvssslsisvpslatsalnprftfpshltqsgesaekngrprkqtsa-paepfsss 768

Qy 412 QRRVKF-----INMGLMED--PMKVYKDRQFMNVWTDHEKEIFKDKF-----452

Db 769 sptplfpwftpgsqtergrnkdapeelskdr-----dadksvekdksrerdrereke 821

|    |                                                                 |     |
|----|-----------------------------------------------------------------|-----|
| Q1 | QYHFRNFGLLRRKSVFDCVALLIILLANNE                                  | 433 |
|    | :   :   :                                                       |     |
| Pb | 822 nkresrkekrkkaseigsssa yvngvrsvkekvcedvatssakkatarkkssshdsat | 881 |

|     |                                                                 |     |
|-----|-----------------------------------------------------------------|-----|
| 022 | UNRESIKENIAYSGSEIYSSAIPVGIIVSKKVVJGUAUWALSSSSAKKALGIKKKSSSSUOGL | 886 |
| 09  | -----NYKALVRRNYSKKRRGRNOOIARPSOEKVEEK-----                      | 517 |

|    |     |                                                             |       |    |
|----|-----|-------------------------------------------------------------|-------|----|
| QY | 460 | -----NINATLVKRNKGNKGNQQLARPSQVEEREA                         | ----- | 31 |
|    |     | : : :                                                       | : :   |    |
| Pb | 882 | ditsvtlqdtavvktkilkikgkanlektndlq-ptapslekektlclstpsstvykhs | 940   |    |

004 ulesvcsyucclavakkkrrkkkgygurekknunuy pnapstiekekrcrscrspsssslvkms 94















PT - which inhibits and/or promotes the activation of cAMP and mitogen  
 responsive genes

XX Claim 1; Page 32-39; 46pp; English.

CC Identification of a cpd. (1) which inhibits and/or promotes  
 activation of cAMP and mitogen responsive genes comprises monitoring  
 expression of a reporter in response to (1), relative to expression  
 of reporter in the absence of (1), where exposure of (1) is  
 performed in the presence of (1) a signal dependent transcription;  
 CC (2) a polypeptide comprising at least amino acid residues 461-661 of  
 this sequence (AAR79054); and (3) a reporter construct comprising a  
 CC reporter gene under the control of a signal dependent transcription  
 factor. The method can be used to identify a compound which has  
 CC the binding and/or activation properties of CREB binding protein or  
 CC the transcription activation characteristic of a signal dependent  
 transcription factor.

XX Sequence 2441 AA;

Query Match 2.48; Score 309.5; DB 16; Length 2441;  
 Best Local Similarity 18.7%; Pred. No. 2.4e-08;  
 Matches 495; Conservative 304; Mismatches 992; Indels 859; Gaps 122;

QY 7 PPNQAFSTEQSRYPHPSVQYTFNTRHQEFVAVDYRSHLEVSVQASOLQQOQOQOQ-- 64  
 DB 121 plngdst-----pnlpka-----astsgptppasqalnpaqkqvg 159  
 QY 65 LRRRPSLLSEPHGSDRQERTSYEPHPG--PSPVDHDSLEKRRPLEQVSDSHF--- 119  
 DB 160 lvtspatsqtgpgic-----mnanfnqthpgllnnsghslmnaqagqgqamngslgaa 215  
 QY 120 QRVSAVLPLVHPLPEGLRASADAKK---DPAPGGKHEA----- 155  
 DB 216 grgrgagmpapamqagatvlaetlqvspqmag-haglnaqaggmktmgmtgttsp 274  
 QY 156 ---PSSPTSGQPCGDQNASPSKLSKEELIQSDMRDVRDIAKVEQIILKLLKQKQOQLEEE 212  
 DB 275 fgqfsgtggqmg-atgvnqlaskqsmvnslpaftdiknts---vttvpmmsqlqts 330  
 QY 213 AAKPPEPKVPSPVPEQKHSIVQ-----IYDENRKAEEAHKIFELGLBKVLPVLYNQ 268  
 DB 331 vgiptqaiatgptadpekrklqqlvlllhakcqrqean---gevrcacslp----- 382  
 QY 269 PSDTKVYHENIKTNQVMKLLILFFKRRNHARKOREQKICO-----RYDQLMEAWEK--K 321  
 DB 383 -----hcrf--mknvl-----nmthcqapkcacqvahcaserqliishwkncr 423  
 QY 322 VDRIENNPRRKAKESKREYYEKFPEIRKOREQOERF-----QRVGORGAGLSATI 373  
 DB 424 hdcpcvclpknasd-----krnqqtllgspasglqntigsvgagqgnat 467  
 QY 374 ARSEHETSEIIDGLSEQB-----NNEKOMRQLSVIP---PMFDAEORRVKFNNM 421  
 DB 468 slsn---pnpidpsmggrayaalpymnqptqlqpqvpgqgpaqpbahq---qmrtln 521  
 QY 422 GLMEDPMKV-----YKQDFMNVTDHEKEI---FKDKFIQHPKNFGLIASYLERKSVDP 473  
 DB 522 algnpmvsvpaggittddqppnlisesalptslgatpmlmdngsnsgnigsl---stip- 577  
 QY 474 CVLYXYLTKNENYKALVRNRYKRRNRNQOIARPSOEKVEEKEE-----DKAEKTEK 527  
 DB 578 -----taappstgvrkgwhehvtqdlshlvhkvqaiftpdpaaakdr 623  
 QY 528 KEEE-----KKDEEKDEKESKENTKE--KDKTIGTAEETEEERQATPRGRKTANSQGR 580  
 DB 624 rmenlvayakkevgydesansrdeyhlhlaeklykqlelekr-----rtrlhkqg- 676  
 QY 581 RKGRITRSMWNEAASAAAAATEEPP-----PPLPPPPPEPISTPEVETSRWTEEEVEV 635  
 DB 677 -----ilgnqpalpasgqppvipagvrbpnpplpvpnmqvsqgmnsfnpmsl 728

QY 636 AKKGLVEHGRNWAIAKM---VGTKSEAQCKNFYFNKRRHNLNLLQOQKOK-TSRKPR 691  
 DB 729 gnvqlpqagmgpraaspmhsvqmnsmasvpgmaismrmpqppnmmtgthannimaqapt 788  
 QY 692 EERDVSOCE-----SVASTVSAQEDDEDEASNEEE-----NPEDESEAV 731  
 DB 789 qnqflpqnqfssgamsvsmgmgpaaagqvsqgqepgaalpnlmnapqasqipc- 847  
 QY 732 KPSEDSP-----ENATSRGNT--EPAVELEPTETAPSTSPSLAVPST 772  
 DB 848 ppvtqslphtpppastaagmpsiqhptapgmtpqpapaptqstpvssgqgtpt-ptpgs 906  
 QY 773 KPABDESVEVDNDISGAETAQMDVQOQHSAGEGSCVCDP-----PPAT-- 817  
 DB 907 vps---aaqtqstvtvaaaqav--tpqptqvpvppsvatpqssqqqptpvhtqppgtpl 962  
 QY 818 --KADSVDEVRVPENHASKVGEVDNTKERDLDRASEKVEPRDEDLVVAQOINAORPEQS 875  
 DB 963 sqaasidnrvtpstvtstaetsqqpqpdpvmlmlemktevqtd-----aepepte 1013  
 QY 876 DNDSSATCSADEVDGEPERQRMFMDSKPSLLNPTGSLVSSPLKPNPLDLPOLOHRAA 935  
 DB 1014 skgeprsemmeedlqgssqvkeetd-----tteqksepmveveekkpvk 1057  
 QY 936 VIPPMVSTPCNIPITGPVSGYALYQRHIKAMHESALLEEQORQEQIDLECRS-STSPC 994  
 DB 1058 v-----eakeeeensndtasqstps 1079  
 QY 995 GTSK---SPNREWEVLOPAPHQLITNLPBGVRLTPTRPPPLIPSSKTTVASEKPSF 1051  
 DB 1080 qprkikfpeelrqlmptlealryqdpes--lprgqvdpqllgipdyfdivknpmdls 1137  
 QY 1052 IMGGSISOG---TPGTY-----LTSHQASVTOETPK-----PSVGS 1086  
 DB 1138 tikrkidtygqepqvdydvrlmfnawlynrktsvyfkscklaevfegeidpvmqsi 1197  
 QY 1087 SLGLPROQESAKSATLPYIKOEESPRSQNSQOPEGLLVRQAQHEGVVVGTA-GAIQEGSIT 1145  
 DB 1198 gycgrgyefspqtlccyqkqlctipr-----daayysyqnrhyfcgkctelqgenvt 1251  
 QY 1146 RGTPTSISVESIPSLRGSITQGTIPALPOTGIPTEALVKGSIIRMPIDEDSSPEKREAA 1205  
 DB 1252 lg-----ddpsqpttiskdqfekknndtl---dpepfvdckecg 1288  
 QY 1206 SKGHVIVEGSGHILSYDNI-----KNAREGTRSPRTAHEISLRSYSEVEG- 1253  
 DB 1289 rkmhqic-----vlhydiwpsgvfcdncikktgrprkenkfakrlqttrlnghedr 1342  
 QY 1254 IKOGMSMRSPVSAPLEGLICRALPRGSPHSD-----LKERTVLGSGTMOQTTPRATT 1305  
 DB 1343 vnkflrrqnheag-----evfvrivassdktevkvpgmksrfvdsgemesefpyrtk 1395  
 QY 1306 ESFEDGLKYPKQIKRESPPIRAFEGATKGPYDGITIKEMGRSIIHE---IPQDILT 1361  
 DB 1396 alf-----afe-----eidgv-dvcffgmhvqdtaliaphq----- 1425  
 QY 1362 QESRKTPEVWQSTRPIIEGS--ISOGTPIKFDN-NSGOSAIKHN-----VKSLLI--- 1407  
 DB 1426 -----lqgcviysldsihfrprclrtavvheillgyleyvkllyvt 1469  
 QY 1408 -----TGPSKLSRGMPPLEIVPENIKVVERGK-----YEDV--KAGE 1442  
 DB 1470 ahiwacppsegddyfchcpddqkpkprkrlqewykkmlkdkaferliindykdifkane 1529  
 QY 1443 TVRSRHTSVV-----SSGSPVLRLTLHEAPKAQLSGFIYDDTSARITPVSYQ-----NT 1491  
 DB 1530 ---drlltsakelpyfeqdfvnpvleesikeleqegeeeerkkeestaasetpsgsgdskna 1586  
 QY 1492 MSRGSPMMNRT-SDVT-----IPPNKSTNHERKSTLTPTORES-----IPA 1531  
 DB 1587 kknknktknksisrankkksmpnvnndlsqklyatmekhkevfvfvlhlagpvlst 1646  
 QY 1532 KSPVPVGDVPVVS-----HSPF-----DPHHRGSTAGEVYWS----- 1562



Db 424 hdpvcplnknsd-----krnqdtllgspasgiqntigsvagqgnat 467  
Qy 374 ARSEHEISEIIDLSEOE-----NNEKOMROLSIP-----PMFDAEORRRKFIMN 421  
Db 468 slsn---pnpldpssmqrayaalglpymnqdtqlqpyvpgqpaqppahq---qmrtiln 521  
Qy 422 GLMEDPMKV-----YKDRQFMVWMDHKEI---FKDRFIQHPNFGLIASLYERKSVPD 473  
Db 522 algnpmsvpaggittdqgpnllisalsalptslgatnplmdngsngnigs1---stip- 577  
Qy 474 CVLYXYLLTKKENYKALVRNRYKRRGRNOQIARPSQEKVEEKEE-----DKAEKTEK 527  
Db 578 -----taappstgvrkvwhehvtqdlrshlvhklvqalfptdpaaalkdr 623  
Qy 528 KEE-----KKDEEKDEKDESKENTKE--KDKIDGTAEETEEREQAPRGKKTANSOGR 580  
Db 624 rmenlvayakkvegmyesansrdoyhllaeklykikqeleer-----rtrlhkgg- 676  
Qy 581 RKGRITRSMTEAAAAAATAATEPP-----PPLPPPPPEPISTEPVETSRWTEEMEVEV 635  
Db 677 -----ilgnqpalpasgagppvippaqsvrppngplp1pvnrmqvsqgmnsfnpmsl 728  
Qy 636 AKKGLVEHGRNWAATAKM---VGTKEBAQCKNFYENYKRRHNDLNLQOHKQK-TSRKPR 691  
Db 729 gnvqlpqsapmgpraaspmnhsvqmnsmasvpgmaispmpqpnpnmngthannimaqapt 788  
Qy 692 EERDVSCOE-----SVASTVSAQEDIEDIEASNEEE-----NPEDSEVEAV 731  
Db 789 qnqlfpqnfssagamsvnsvmgpaagagvsgqgegaalpnpplmlapqasqlpc- 847  
Qy 732 KPSEDSP-----ENATSRGNT--EPAVELEPTTETAPSTSPSLAVPST 772  
Db 848 ppvtqslphtpppaastaagmpsiqlhtcapgmtppqpaaptqstpvssggtpt-ptpgs 906  
Qy 773 KPAEDESVEVQVNDISIAETAEQMDVDQOEHSABEGSVCDP-----PPAT-- 817  
Db 907 vps---aaqtstqtpvqaaagqv--tpqptpvqpsvatpqssqgqtpvhtppgtpl 962  
Qy 818 --KADSVDEVRVBNHASKVEGDNTKERDLDRASEKVEPDEDLVAAQINAOGRPEPOS 875  
Db 963 sqaaasidnrvtptstvtasetsqgppdvpmlemktevtdd-----aepepte 1013  
Qy 876 DNDSSATCSADEDVDGEPEPERQMPDMSKPLLNPNTGTSILVSPKLPNPLDLQLOHRAA 935  
Db 1014 skgeprsemneedlqgssqvkeetd-----tteqksepmveeekbevk 1057  
Qy 936 VIPPMVSTCPNIPGTVPVSGYALYQRIHAKHESALLEEORQOEQIDLECRS-SRSPC 994  
Db 1058 v-----eakeeeenssndtasqstps 1079  
Qy 995 GTSK---SPNREWEVLOPAPHQITLNLPEGVRLPTRTRPPPLIPSSKTTVASEKPSF 1051  
Db 1080 qprkikfpeelrqlmptlealryqdes--lprqvdvqlligidyfdiivknpmcls 1137  
Qy 1052 IMGSIQSG---TPGTY-----LTSHQASYTOETPK-----PSVGS1 1086  
Db 1138 tikrkltdgqyqevqyvdvrlmfnawlrynrktsrvyfkcsklaevfeqeidpvmqsl 1197  
Qy 1087 SLGLPQOESAKSATLPIKQEEFSPRSQNSQOPELLVRAQHEGVVRGTA-GATQESGIT 1145  
Db 1198 gyccgrkyefspqtlccyqgkqlctipr-----daayysyqnrhfcgkcfteiqgenvt 1251  
Qy 1146 RGTPTSKISVESIPSLRGSITQGTTPALPQTGIPTEALVKGSI SMPDIETSSPEKGREAA 1205  
Db 1252 lg-----dpsqpttiskdqfekknndtl---dpepfvdckecg 1288  
Qy 1206 SKGHVIVEGKSGHILSYDNI-----KNAREGTRSPRTHAIEILSKRSYESVEGN----- 1253  
Db 1289 rkmhqc-----vlhydiilwpsgfvcdnclktgprkenkfsakrlqtrlguhledr 1342  
Qy 1254 IKQGSMSRSPVAPLEGLICRALPRGSPHSD-----LKERTVLSGSIWQGTTPRAT 1305  
Db 1343 vnkflrrqnhpeag-----evfvrvvassdktvckpkmksrfvdgmsesfprtk 1395

Qy 1306 ESFEDGLKYPKQIKRESPPIRAFEGAITKGPYDGIITTIKEMGRSIHE-----IPRODILT 1361  
Db 1396 alf-----afe-----eidgv-dvcffgmhvqdtaliabhq----- 1425  
Qy 1362 QESRKTPEVWOSTRIIEGS--ISQGTPIKFDN-NSGOSAIKHN-----VKSLI--- 1407  
Db 1426 -----lqgcvyisylsifhfrprclrtavyheillqylevkklyvt 1469  
Qy 1408 -----TGPSKLSRCMPPLEIVPENIKVVERCK-----YEDV--KAGE 1442  
Db 1470 ahiwacppsegddyfihchppdqkpkpkrilqewykkmlkdkaferiindykdifkgane 1529  
Qy 1443 TVSRHRHSVW-----SSGSVLRLSTLHEAPKAQLSPGIYDDTSARSTPVSQO---NT 1491  
Db 1530 ---drltsakelpyfedgfwpnvleesikeleqeeerkeesteasetpagsqdsna 1586  
Qy 1492 MSRGSPMMNRT-SDVT-----IPPNKSTNHERKSTLTPTQRES-----IPA 1531  
Db 1587 kknkktknkssistrankkkmpnvnslsgklyatmekhkevffvihlhagpvist 1646  
Qy 1532 KSPVPGVDPVVS-----HSPF-----DPHHRGSTAGEVYS----- 1562  
Db 1647 qpivdpopllscldmdgrdaifltardkhwefslrrskwsticmlvelhtqgdrfvy 1706  
Qy 1563 -----HLPTQ-----LDPAMPFHRAL-----DPAATAAYL 1586  
Db 1707 tcneckhhveirwhctvcedydlcincyntskthkmvkwglgiddegssgeqpskspq 1766  
Qy 1587 FOROLS-----PPPGYPSQYQLYAM 1606  
Db 1767 esrrlsiqrciqslsvhacqcrnancsipsccqkmkrvvqhtkgckrtnggpcvckqlial 1826  
Qy 1607 -----ENTRQTILNDYITSQOMQ-----VNLRPDVARGL-SP 1637  
Db 1827 ccyhakhcquenkcvpfclnlkhnvrqgqlchicqagqlmrmatntrnvpqslpsp 1886  
Qy 1638 REQPLGLPYPATRGIIIDLTNNPPITLVPHPGCTSTP-PMORI-TVIPGTQITFFP----- 1690  
Db 1887 tsapggtp-----tqqpstqtpqppaqppavpnmnspagfpnvartqptvlsva 1936  
Qy 1691 -RPNYASMSGPHPTHAAAAAERERERERERERERERERERERERERERERERERER 1749  
Db 1937 gkptcnypappppaqppaaaveaaraqieraaqqghlyr---aninngmpbgdgmgtpg 1993  
Qy 1750 SHGYVRSPSPSVRTOEHWLOQRPVSFQGTNGTSTVITPLDPTAQIRIMPLPAGGSISQGL 1809  
Db 1994 sq-----mtpvglnvprpnqvgpvmssm-----ppgwgqgapiqqqp--mpgm 2036  
Qy 1810 PASRYNTAADALAALVDAASAPOMDVSKTESKHEAARLEENLRSSAAVSEOOOLEOK 1869  
Db 2037 p-----rpvmsnqqaavaagvprmpnvqprnsipsa-lqdlrltlkspsspqgq--qq 2086  
Qy 1870 TLEVERSVQCLYTSFAP-----PSGKPP--H-----SSV 1898  
Db 2087 vlnilksnqilm---aafikqrtakyvanqgmqpqpglqsgqgmqpqgmhqqpslqnl 2143  
Qy 1899 VYSEAGDKK-GPPPKSYEEELRTRGKTTITAAFNID-----VITROI- 1941  
Db 2144 namqagvprpqvpppqbamggnpqgq---alnimpghnmpnmnpqyrenvrrqll 2199  
Qy 1942 --ASDKDARERGSSQSSSSSLSS-----HRYETPDAIEVISPAS-SPAPPOEKLOTQY 1993  
Db 2200 qhqgqqqqqqqqqqqqqqnsaslagmaghsqfqqp-----gpggyapamqqrmaqhl 2253  
Qy 1994 P-----EYVKAQO-----AENDPTRO---YEGPLHHYRPOQESPSPQOQLPP 2032  
Db 2254 pigssmqmaapmgqlqgmqqpqlgadstpnigalqqrllqgqgmkgqigsqqpnpnm 2313  
Qy 2033 SSQAEQMGQVPRTHRLITLADHI--COIITQDFARNQVSSOTPOQPTSTFQNSPSALVS 2090  
Db 2314 spqghmsigpq-----ashlpggqlats--lsndvrspapvqspbrq-----s 2355









```

Db 2262 qmaapmgqlgmgqpglqadstpnliqqalqqriliqqqmkqigspgqpnmpspqqhmls 2321
QY 2041 QVPRTHRLITLADHI--CQIITQDFARNQVSSQTPQQPPTSTFONSPSALYSTPVRTKTS 2098
Db 2322 gqpq-----ashlpqqiats--lsnqvrsapvqgsprpq-----sqpphssps 2363
QY 2099 NRYSPESOAQSVHH-----QRGSRVSPENLYDKSRGSRPGKS-----PERSHVS 2143
Db 2364 priqpq---psphhvspqtgtphpglavtmasmdqghlgnpedsamlpqintpnrsals 2420
QY 2144 SE 2145
Db 2421 se 2422

```

Search completed: September 8, 2001, 14:42:18  
Job time: 480 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 14:35:13 ; Search time 23.22 Seconds  
(without alignments)  
2163.671 Million cell updates/sec

Title: US-09-522-753-11

Perfect score: 12643

Sequence: 1 MSSSGYPNQGAFTSQSRY.....EREPAPLLSAQYETLSDSD 2440

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description      |
|------------|--------|-------------|--------|----|------------------|
| 1          | 2885.5 | 22.8        | 619    | 2  | US-08-372-652-5  |
| 2          | 2885.5 | 22.8        | 619    | 5  | PCT-US93-16311-5 |
| 3          | 368    | 2.9         | 2843   | 1  | US-07-741-940-2  |
| 4          | 368    | 2.9         | 2843   | 1  | US-08-289-548A-2 |
| 5          | 368    | 2.9         | 2843   | 1  | US-08-452-654-2  |
| 6          | 368    | 2.9         | 2843   | 2  | US-08-370-235A-2 |
| 7          | 365    | 2.9         | 2842   | 1  | US-07-741-940-7  |
| 8          | 365    | 2.9         | 2842   | 1  | US-08-289-548A-7 |
| 9          | 365    | 2.9         | 2842   | 1  | US-08-452-654-7  |
| 10         | 365    | 2.9         | 2843   | 1  | US-08-452-655B-2 |
| 11         | 365    | 2.9         | 2843   | 1  | US-08-452-655B-7 |
| 12         | 365    | 2.9         | 2843   | 3  | US-08-450-582-2  |
| 13         | 365    | 2.9         | 2843   | 3  | US-08-450-582-7  |
| 14         | 365    | 2.9         | 2973   | 2  | US-08-821-355A-7 |
| 15         | 365    | 2.9         | 2973   | 2  | US-09-003-687A-7 |
| 16         | 365    | 2.9         | 2973   | 4  | US-09-136-605-7  |
| 17         | 329    | 2.6         | 3969   | 4  | US-08-061-376-5  |
| 18         | 323.5  | 2.6         | 1780   | 1  | US-08-769-309A-5 |
| 19         | 323.5  | 2.6         | 1780   | 3  | US-08-994-570-5  |
| 20         | 309.5  | 2.4         | 2441   | 1  | US-08-194-468-2  |
| 21         | 309.5  | 2.4         | 2441   | 3  | US-08-961-773-2  |
| 22         | 309    | 2.4         | 1719   | 2  | US-08-459-568-4  |
| 23         | 309    | 2.4         | 1719   | 3  | US-08-399-411-4  |
| 24         | 309    | 2.4         | 1719   | 3  | US-08-516-859A-4 |
| 25         | 294.5  | 2.3         | 2035   | 1  | US-08-046-585-5  |
| 26         | 294.5  | 2.3         | 2035   | 1  | US-08-393-703-5  |
| 27         | 294.5  | 2.3         | 2035   | 5  | PCT-US93-11721-5 |

Sequence 4, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 94, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 4, Appli  
Patent No. 5180808  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 23, Appli  
Sequence 1, Appli

28 286.5 2.3 2101 1 US-08-195-487-4  
29 286.5 2.3 2101 5 PCT-US93-06160-4  
30 285.5 2.3 2101 1 US-08-466-390-4  
31 285.5 2.3 2101 1 US-08-470-950-4  
32 285.5 2.3 2101 1 US-08-467-781-4  
33 285.5 2.3 2101 2 US-08-483-924-4  
34 284 2.2 1898 1 US-08-056-200-94  
35 284 2.2 1898 2 US-08-800-644-94  
36 284 2.2 2414 1 US-08-227-536-2  
37 284 2.2 2414 5 PCT-US95-04682-2  
38 276.5 2.2 1618 1 US-07-853-913-4  
39 275 2.2 2409 6 5180808-2  
40 274.5 2.2 1706 2 US-08-459-568-2  
41 274.5 2.2 1706 3 US-08-399-411-2  
42 274.5 2.2 1706 3 US-08-516-859A-2  
43 270.5 2.1 1805 1 US-07-853-913-2  
44 258 2.0 1185 4 US-09-041-886-23  
45 256.5 2.0 1829 4 US-09-157-420-1

#### ALIGNMENTS

RESULT 1  
US-08-372-652-5  
; Sequence 5, Application US/08372652  
; Patent No. 5932699  
; GENERAL INFORMATION:  
; APPLICANT: Moore, David  
; APPLICANT: Seol, Wongi  
; APPLICANT: Choi, Hwang-Sik  
; TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING  
; POLYPEPTIDES AND RELATED MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street, Suite 3100  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/372,652  
; FILING DATE: 13-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00786/246001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 619 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
US-08-372-652-5

Query Match 22.8%; Score 2885.5; DB 2; Length 619;  
Best Local Similarity 90.1%; Pred. No. 2.4e-171;  
Matches 562; Conservative 27; Mismatches 28; Indels 7; Gaps 3;

QY 1819 DATAALYDAASAPQMDVSKTKESKHEAARLEENLRSAVSEQQOLEQKTKLEVKRSV 1878  
|||||

Db 1 DALAALVDAASAPQMDVSKTKESKHEAARLEENLRSSAAVSEQQOLEQKNLEVERSV 60  
QY 1879 QCLYTSSAFPSGKQPQSHSSVYVYSEAGKDKGPPPKSRVEEELRTRGKTTITAANFIDVIT 1938  
Db 61 QCVCTSSALPSGKAQPHASVYVYSEAGKDKGPPPKSRVEEELRTRGKTTITAANFIDVIT 120  
QY 1939 QTIASDKDAREGQSSDSSSLSSHYETPSDAIEVISPASSPAPQEKLTQYQPEVVK 1998  
Db 121 QTIASDKDAREGQSSDSSSLSSHYETASDAIEVISPASSPAPQEKLTQYQPEVVK 180  
QY 1999 ANQENPTROYEGPLHYRQEQSPSPQOQ--LPPSSQAGMGQVPRTHRLITLADHIC 2056  
Db 181 ANQENESTROYEGPLHYRQEQSPSPQOQPLPPSSQSEGMQVPRTHRLITLADHIC 240  
QY 2057 QITQDFARNQVSSQTPQPTSTFQSPSALVSTPVRTKTSNRYSPESQASVHHORPG 2116  
Db 241 QITQDFARNQV---PSQPSSTFTQSPSALSTPVRTKTSNRYSPESQASVHHORPG 296  
QY 2117 SRVSPENLVKSRGSRGKSPERSHVSSEPEYPIPPQVPPVHHQKQSLLSQSGAEPA 2176  
Db 297 PRVSPENLVKSRGSRGKSPERSHIPSEPEYPIPPQPAVHKEKQDSMLLSQSGVDP 356  
QY 2177 EORNDARSQGISYLPFFTKLENTSPMVKSKKOEIRKLNSSGGGSDMAAOPGTEIF 2336  
Db 357 EQRSDRSRSGISYLPFFTKLESTSPMVKSKKOEIRKLNSSGGGSDMAAOPGTEIF 416  
QY 2237 NLPVTTSGVSSSRGSHFADPNLGLDIIIRKALMGSDFKVEDHGVMSQVHHORPG 2296  
Db 417 NLPVTTSGVSSSRGSHFADPNLGLDIIIRKALMGSDFKVEDHGVMSHPVIMPGS 476  
QY 2297 ANTSVVTSGTTRREGDPSHSGGVCCKPKLSKNSRKSPIPGQYLGTERPSSVSV 2356  
Db 477 ASTSVVTSSEARRDEGEPSPHA-GVCKPKLINKNSRKSPIPGQYLGTERPSSVSV 535  
QY 2357 HSEGDYHROTPGAWEDRPSSTGSTQPPYNPLTIRMLSSPTPTIACAPSVAQAAPHQ 2416  
Db 2417 NRIWEREPAPLLSAQYETLSDSD 2440  
Db 596 NRIWEREPAPLLSAQYETLSDSD 619

## RESULT 2

PCT-US95-16311-5  
; Sequence 5, Application PC/TUS9516311  
; GENERAL INFORMATION:  
; APPLICANT: Moore, David  
; APPLICANT: Seoul, Wongi  
; APPLICANT: Choi, Hwang-Sik  
; TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING  
; TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street, Suite 3100  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/16311  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/372,652  
; FILING DATE: 13-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.

; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00786/246001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 619 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
PCT-US95-16311-5

Query Match 22.8%; Score 2885.5; DB 5; Length 619;  
Best Local Similarity 90.1%; Pred. No. 2.4e-171;  
Matches 562; Conservative 27; Mismatches 28; Indels 7; Gaps 3;  
QY 1819 DALAALVDAASAPQMDVSKTKESKHEAARLEENLRSSAAVSEQQOLEQKNLEVERSV 1878  
Db 1 DALAALVDAASAPQMDVSKTKESKHEAARLEENLRSSAAVSEQQOLEQKNLEVERSV 60  
QY 1879 QCLYTSSAFPSGKQPQSHSSVYVYSEAGKDKGPPPKSRVEEELRTRGKTTITAANFIDVIT 1938  
Db 61 QCVCTSSALPSGKAQPHASVYVYSEAGKDKGPPPKSRVEEELRTRGKTTITAANFIDVIT 120  
QY 1939 QTIASDKDAREGQSSDSSSLSSHYETPSDAIEVISPASSPAPQEKLTQYQPEVVK 1998  
Db 121 QTIASDKDAREGQSSDSSSLSSHYETASDAIEVISPASSPAPQEKLTQYQPEVVK 180  
QY 1999 ANQENPTROYEGPLHYRQEQSPSPQOQ--LPPSSQAGMGQVPRTHRLITLADHIC 2056  
Db 181 ANQENESTROYEGPLHYRQEQSPSPQOQPLPPSSQSEGMQVPRTHRLITLADHIC 240  
QY 2057 QITQDFARNQVSSQTPQPTSTFQSPSALVSTPVRTKTSNRYSPESQASVHHORPG 2116  
Db 241 QITQDFARNQV---PSQPSSTFTQSPSALSTPVRTKTSNRYSPESQASVHHORPG 296  
QY 2117 SRVSPENLVKSRGSRGKSPERSHVSSEPEYPIPPQVPPVHHQKQSLLSQSGAEPA 2176  
Db 297 PRVSPENLVKSRGSRGKSPERSHIPSEPEYPIPPQPAVHKEKQDSMLLSQSGVDP 356  
QY 2177 EORNDARSQGISYLPFFTKLENTSPMVKSKKOEIRKLNSSGGGSDMAAOPGTEIF 2236  
Db 357 EQRSDRSRSGISYLPFFTKLESTSPMVKSKKOEIRKLNSSGGGSDMAAOPGTEIF 416  
QY 2237 NLPVTTSGVSSSRGSHFADPNLGLDIIIRKALMGSDFKVEDHGVMSQVHHORPG 2296  
Db 417 NLPVTTSGVSSSRGSHFADPNLGLDIIIRKALMGSDFKVEDHGVMSHPVIMPGS 476  
QY 2297 ANTSVVTSGTTRREGDPSHSGGVCCKPKLSKNSRKSPIPGQYLGTERPSSVSV 2356  
Db 477 ASTSVVTSSEARRDEGEPSPHA-GVCKPKLINKNSRKSPIPGQYLGTERPSSVSV 535  
QY 2357 HSEGDYHROTPGAWEDRPSSTGSTQPPYNPLTIRMLSSPTPTIACAPSVAQAAPHQ 2416  
Db 536 HSEGDYHROTPGAWEDRPSSTGSTQPPYNPLTIRMLSSPTPTIACAPSVAQAAPHQ 595  
QY 2417 NRIWEREPAPLLSAQYETLSDSD 2440  
Db 596 NRIWEREPAPLLSAQYETLSDSD 619  
RESULT 3  
US-07-741-940-2  
; Sequence 2, Application US/07741940  
; Patent No. 5352775  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/741,940  
FILING DATE: 19920109  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.035574  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-741-940-2

Query Match 2.9%; Score 368; DB 1; Length 2843;  
Best Local Similarity 19.4%; Pred. No. 4.3e-14;  
Matches 418; Conservative 281; Mismatches 797; Indels 664; Gaps 98;

Qy 347 PEIRKQEQE-RFQRVQRCAGLSATIRSEH-----EISEIIDGLSQENNEK 395  
Db 981 PSIESYEDDESKEFCYQVPADLAHKIHSANHMDDNDGELDTPTNYSLSKYSDQOLNSGR 1040

Qy 396 QMRQLSVIPMMFDAQRRVKFINNGLMEDPMKVYKDRQFMNVTWTHE--KEIFKDKFI 453  
Db 1041 Q-----SPQNERWARPKHIIEDIEIKQSEQRQSRNOSTYYPVYTESTDDKHL 1087

Qy 454 QHPKNFGLIASYLERKSVPCDVLVYYTKKNENYKALVRRNRYGKRRGNQOIARPSQEK 513  
Db 1088 KFQPHFG-----QQECVSPY-----RSRGANGSETNRVGNHGINQVVSQSLQOE- 1132

Qy 514 VEEKEDEKAETK--EEKKDEKEDKSKENTKDKIDGTAEETEEREQAPRG 571  
Db 1133 -DDYEDDKPTWYERYSEEEQHEEERTNTYSIKYN-EEKRHVDQPIDYSUKYATDIPSS 1190

Qy 572 R-----KTANSQRRKGRITRSTNEAAAS-----AAAAATEEPPPLPPP 614  
Db 1191 QKQSFSEKSSGSSGOSTEHHSSSENTSTPSSNAKRONQLHPSSAQSRSQPKQAATCK 1250

Qy 615 PEPITSTPEVET-----SRWT-----EEMEVAKKGLVEHGRNWAIAKM--- 653  
Db 1251 VSSINQETIQTVCYEDTPICFSRCSLSLSAEDIEGCTOTQOADSANTLQIAEIKGK 1310

Qy 654 VGTKSE-----AQCKNFYNYKRRHNLNLLQOHK-----QKTSR 688  
Db 1311 IGRSAEDPVSEVPASQHPRTKSRSLQGSLSSESARHKAVFPFGSAKSPSKSGAQTPK 1370

Qy 689 KPREE-----RDVQCESVASTVSAQDEDEIAGNEEBENPEDSEVAVKPS--DSPEN 740  
Db 1371 SPPEHYVQETPLMFRCITSV--SSLDSFESRSTASSVQSGPCSGMVGSIISPSDLDFDSQO 1429

Qy 741 A--TSRGNTPEPAVELEPTTAPSTSPSLAVPSTKPADESVEVETOVNDSISAETAQMDV 798  
Db 1430 TMPSPRSKTPPP--PPQTAQTKREVPKNAKPTAEKRESGPKQAANAQV--RVQVLQDA 1485

Qy 799 DQOEHSAEE---GSCV-----DPPPATKADSVDEYR-----VPEN-HASKVEGDN 840  
Db 1486 DTLHLFATESTPDGFCSSSISALSLEDFIOK-----DVELRIMPVQVQENDNGNETESEQ 1541

Qy 841 TKE--RDLDRASEKVEPRDELDLVAQINAOERPEQSDNDSSATCSADEVD----- 890  
Db 1542 PKESNEQEKAEKTIIDSEKDLL-----DSD-----DDDIIECEIIS 1581

Qy 891 -----GEPERQRMF-----PMSDKPSLLNPTGSIILVSSPLKPNPLDLPLQOLHRAAVI 937  
Db 1582 AMPTKSSRKGKKAQASKLPPVARKPSQL--PVYKLLPSQ-----NRL---QPQKHVSFT 1633

Qy 938 P----PWVSC---TPCNTPIGTVPVSGYAL-----YQRHKAHME 969  
Db 1634 PGDDMPRYCYVEGTPINFSTATSLDLTIESPPNELAAGEVYRGGAQSGEFKRDITIPE 1693

Qy 970 SAL-----LEEQRORQEQIDLECRSSSTSPCGTSKSPNREVL----- 1007  
Db 1694 GRSTDEAOGGKTSSVTIPELDDNKAEGDILAEICINSAMPKSKHPPRVKIMDOVOQA 1753

Qy 1008 ---QAPHQLITNLPPEGVRLPTTRPPPLI-----P 1038  
Db 1754 SASSSAPNK---NQLDGKKKPTSPVKPIQONTYRVRKNADSKNNLNAERFSDNKK 1810

Qy 1039 SSKTTVASEKPSF-----IMGGSIS-----QGTPTGYLTSHNOASTQETP 1079  
Db 1811 SKQNLKNNKDFENDKLPNNEDRVGSEAFSPHHTPIEGTP--YCFSRND----- 1860

Qy 1080 KPSVGSISLGLPROQESAKSATLVYKQEEFSPRSONSQPEGLLVRAQHEGVVVRGTAGAI 1139  
Db 1861 --SLSLDFDDDDVDLSREKAELKAKENKESEAKVTSHT--LTSNOOSA---NKTQAI 1913

Qy 1140 QEGSITRGTPTSKISVESIPSLRGSITQGTALPOTGPTTEALVK-----G 1185  
Db 1914 AKQPINRGQPKPILOKQS-----TFQSSKIDIPRGAATDEKLQNFATIENTPVCFSHNS 1967

Qy 1186 SISRM-----PIEDSPEKGRGAAAGHVIYBEKSGHI----- 1219  
Db 1968 SLSLSDIDQNNNNKENEPKETETPPDSQGEPSKPOAGYAPKSPHVEDTPVCFSRNSSL 2027

Qy 1220 --LSYDN-----IKNAREGTRSPRTAHEISLRKSYSEVGNIKQGMSRESVPSAPL 1269  
Db 2028 SLSLSDSEDDLQECISSAMPKKKPSRLKGDNEKHSPRNMGGLGEDLTLDLKDQRPD 2087

Qy 1270 EGLICRALPRGSPHDLKERTVLSGIMQGTPTPRATTEFEDGLKYPKQIKRESPIRAFE 1329  
Db 2088 SE---HGLSPDSENFDMKAIQEGANSIVYSLHQAAC-----LSRQASSDSILSK 2139

Qy 1330 GAITKGKPYDGIITTIKEMGRSITHEIPRODILTQESRKTPVYVQSTRPIEGSISOGTPIK 1389  
Db 2140 SGISLGSFP-----HLTPDQEEKPFTSNKGPRILK-----PGEKSTLETKK 2180

Qy 1390 PDNNS--GSAIKHNKSLITGP-----SKLSRGM--PPLPEIPENIKVBERGYEDYKAGET 1443  
Db 2181 IESESGIKGKKVYKSLITGKVRNSNISEISGMKQPLQ---ANMPSISRG----- 2227

Qy 1444 VRSRHTSVVSSGSPVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMWSRGSPMMNTS 1503  
Db 2228 -----RTMIH-----IPGV-----RNSSSSTSPVSKRGP 2251

Qy 1504 DVTIPPNNKSTNHERKSLTPT--QRESIPAK--SPVPGVDPVYVSHSPFDPHHRGSTAGEVYV 1561  
Db 2252 PLKTPASKSPSEGGTATTSPGAKPSVKSELSVPARQTSQIGGSKAPSRSGSR----- 2305

QY 1562 SHLPQDLPAMPFFHRALDPAALAAAYLQRLGPTGPGYQYQYLYAMENTROTILNDYITSQ 1621  
Db 2306 -----DSTPSRPAQQL-----SR 2319  
QY 1622 QMOVNLRPDVA---RGLSPRQPLGLPYPAIRGIIDLINMPTILVPHPGGTSTPPMDRI 1678  
Db 2320 PIOSGRNISIPGRNGISPPNK-----LSQLPRT-SSPSTASTKSGSGKM 2364  
QY 1679 TYIPGTQTFFPPRYNSASMGPHPTHLAAASABERERERE-----KERER 1726  
Db 2365 SY-----TSGRQMSQNLIT---KOTGLSKNASSIPRESASKGLNQMNNGANKVKVEL 2416  
QY 1727 ERIAAASDLRLPGEQGRPGSHGYVRSPSPSVRTQETMLQORPSVFOGTNGTSTVTP 1786  
Db 2417 SRMSSTKSS-----GSESD-----RSERPVLVRQSTFIKEAPS-----2449  
QY 1787 LDPTAQLRIMPLPAGGPISIOGLPASRYNTAADAALVDAASAPQMDVSKTKESKHEA 1846  
Db 2450 ---PTLR-RKLEESASFELS---PSSRPASPTRSQAQTPVLSPLDMSLS-THSSVQAG 2502  
QY 1847 A--RLEENLRSRSAVSEQOOLEOKTLEKRSVQCLYTSSAFPSGKPKOPHSSVYVSEAG 1904  
Db 2503 GWRKLPPNL---SPTIENDGRPAKRHDIA-----SHSESRLPINSGTWKREHS 2552  
QY 1905 KDKGPPPKSRYEEELTRGKTTITAN-----FIDVIITRQIASDKDARERGQSQSD 1956  
Db 2553 KHSSSLP--RVSTWRTGSSSILSASSESEKAKSEKDEKHNVSISGTPKSKENQVSAG 2610  
QY 1957 SSSLSLHRYE--TPSDATEVISPASSAPPOEKLOTQYQPEVVKANQANDPTROVEGLH 2015  
Db 2611 TWRIKIKENEFPTNSTQTVSSGATNGAESKTLIYOMAPAVSK---TEDVWVRIEDCPIN 2667  
QY 2016 HYRPOQESPPQOQLPP-----SSQAEQMGVOPTRHLITLADHI 2055  
Db 2668 NRP---SGRSTGTPPVIVDSYSEKANPNIKDQNDQAKQNGVSPM--RTVGLLENRL 2722  
QY 2056 CQIITQDFARNQVSSQTPOQPT--STFQNSPQALVST---PVRTKTSNRYSPESQAQSV 2110  
Db 2723 TSFIQVD-----APDQKGTEIKPGONNPVPVSETNESPIVERT-----PFSSSSSS 2768  
QY 2111 HQRP---GSRVSPENLVDSRGRPKSPERSHVSSEPEYPISTPPQVPH--EKQDS 2164  
Db 2769 KHSSPGTVAARVTPFNY-----NPSPRKSSADTSARPSQIP--TPVNNNTKKRDS 2818

RESULT 4  
US-08-289-548A-2  
; Sequence 2, Application US/08289548A  
; Patent No. 5648212  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: JOSLYN, GEOFF  
; APPLICANT: KINZLER, KENNETH  
; APPLICANT: MARKHAM, ALEXANDER F.  
; APPLICANT: NAKAMURA, YUSUKE  
; APPLICANT: THLIVERIS, ANDREW  
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
; NUMBER OF SEQUENCES: 102  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Allogretti, LTD  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001-4598  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/289,548A  
; FILING DATE: 12-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 1107,46943  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2843 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-289-548A-2

Query Match 2.9%; Score 368; DB 1; Length 2843;  
Best Local Similarity 19.4%; Pred. No. 4,3e-14;  
Matches 418; Conservative 281; Mismatches 797; Indels 664; Gaps 98;

QY 347 PEIRKOREQE-RQRYQRCAGLSATARSEH-----EISEIIDGLSEQENNEK 395  
Db 981 PSIESYDEDDSKCSYQYPADLAHKTHSANHMDNDGELDTPINYSKYSDEQLNSGR 1040  
QY 396 QMRQLSVIPPMFADQRRVFINNGLMEDPMKYKDRQFMVNTDHE--KEIFKDKFI 453  
Db 1041 Q-----SPQNERWARKHIIIEDEIKQSEQRQNRQSTTPVYTESTDDKHL 1087  
QY 454 QHPKNFGLIASYLERKSPDCVLYLYLTKNENYKALVRRNYKRRGRNQOIAQPSQEK 513  
Db 1088 KFPQHFQ-----QOECVSPY-----RSRGANGSETNRVGNHGINQVNSQLCOE- 1132  
QY 514 VEEKEDEKAEKTEK--EEKKDEEKEDEKEDSKENTKEKIDGTAETETEREQATPRG 571  
Db 1133 -DDYEDDKPTWYSRYSEEEQEEEEERTNYSIKYN--EERHVDQPIDYSKYATFDIPSS 1190  
QY 572 R-----KTANSQRRKRGRITRSMTNEAAAS-----AAAAAATEEPPLPPP 614  
Db 1191 QKQSFESKSSGSGSKTEHSSSENSTSTPSSNAKRNQLHPSSAQSRSGQPKAATCK 1250  
QY 615 PEPITSTEPVET-----SRWT-----EEMEVAKKGLVEHGRNNAIAKM--- 653  
Db 1251 VSSINQETIQYCVEDTPICFSRGSSLSLSSAEDEIGCNOTTOEADSANTLOIAEIKGK 1310  
QY 654 VGTKSE-----AQCKNFYNYKRRHNDLNLQHK-----QKTSR 688  
Db 1311 IGTRSAEDPVSEVPAVSOHPRTKSRLOGSSLSSESARHKAVFPSPGAKSPKSGAQTPK 1370  
QY 689 KPREE-----RDVSOQESVASTVSAQDEDEIASNEENEPEDSEVEAVKPS--DSPEN 740  
Db 1371 SPPEHYVQETPLMESRCTSV--SLDSFESRSIASVQSPGSGVSGIISFSDLPDSGQ 1429  
QY 741 A--TSRGNTPEPAVELEPTTETAPSTPSLAVPSTPKPADESVEVQVNDISISAETAQMDV 798  
Db 1430 TMPPSRSKTPPP---PPQATQTKREVPKNKAPTAEKRESGPKAAVNAVQ--RVQVLPDA 1485  
QY 799 DQOBSAAE---GSVC-----DPPPATKADSDVDVEVR-----VPEN-HASKVEGDN 840  
Db 1486 DTLHFATSTPDGFCSSSSLSLSLDEPFTQK---DVELRIMPVQENDNGNETESEQ 1541  
QY 841 TKE--RDLDRASEKVEPRDEDLVVAQOINAOQPEPQSDNDSSATCSADEVD----- 890  
Db 1542 PKESNQEKEAEKTIIDSEKDL-----DDSD-----DDDEILEECIIS 1581  
QY 891 -----GEPERQRMF-----PMDSKPSLLNPTGSLVSSLPKPNLDLPQHQHRAVI 937  
Db 1582 AMPTKSSRKGGKPAQATASKLPPPVARKPSQL-PVYKLLPSQ---NRL---QPKHVSFT 1633

QY 938 P-----PMVSC---TPCNIPIGTVPYGVAL-----YORIKAMHE 969  
Db 1634 PGDDMPRYCVCEGTPIFNFSATSLSDLTIESPPNELAAGEVGRGAOSGEFEKRDITPTE 1693  
QY 970 SAL-----LEEQROREQIDLECRSTSPCGTSKSPNREWEVL-----1007  
Db 1694 GRSTDEAAGGKTSSTVTIPELDDNKAEEGDILAECINSAMPKGGKSHKPFVKKIMDQVQQA 1753  
QY 1008 ---QAPAHOLLITNLPPEGVRUPPTTPRPPPLI-----P 1038  
Db 1754 SASSAPNK---NOLDGKKKTPSPVPPIQNTYEYTRVRKNADSKNNLNAERVSFNKD 1810  
QY 1039 SSKTTVAASEKPSF-----IMGGSTIS-----OGTPTGYLTSHNOASTOTETP 1079  
Db 1811 SKKQNLANKNSKDFNKLNNEDRYGSAFDPSPHYPIEGTP---YCFSRND-----1860  
QY 1080 KPSVGSISLGLPROESAKSATLPIYIKOEPSPRSONSQPEGLLVRAQHEGVVGRGTAGAI 1139  
Db 1861 --SLSSLDFODDDDVDLSREKAELKAKENKESAKVTSHTE--LTSNQOSA---NKTQAI 1913  
QY 1140 QEGSITRTGTPSKTSVESIPSLRGSITQCTPALPQTGIPTREALVK-----G 1185  
Db 1914 AKQPINRGKPRILQKOS-----TFQSSKDIPDRGAATDEKLQNFALNTVCFSHNS 1967  
QY 1186 SISRM-----PIEDSSPEKRGEEAASKGHVYIEGKSGHI-----1219  
Db 1968 SLSSLDIDQENNNKENEPKETEPDPSQGEPSKPAQSGYAPKSFHVEDTPVCFSRNSSL 2027  
QY 1220 --LSYDN-----INKAREGTRSPRTAHEISLKRYSVEGNIKOGMSMRESVPVSAPL 1269  
Db 2028 SLSIDSEDDLLQICISSAMPKKKKPKSRLKGDNEKSPRNMGGILGEDLTLDLKDIDQRPD 2087  
QY 1270 EGLICRALPGSPHSDLKERTVLGSGIMQCTPRATTESFEDGLAYPKQIKRESPPIRAFE 1329  
Db 2088 SE---HGLSPDSENFDKALQEGANSVSSLHQAAAAAC-----LSRQASDSDSILSLK 2139  
QY 1330 GAITKGPYDGIITIKEMGRSIIHPRQDILITQESRKTPTPEVQVQTRPIEGSISQGTPIK 1389  
Db 2140 SGISLGSFP-----HLTPQEEKPFTSNKGPRILK-----PCKSTLETKK 2180  
QY 1390 PDNNS-QOSAIAKHNVKSLITGP-----SKLSRGM-PPLEIVPENTKVVERGYEDVKAGET 1443  
Db 2181 IESEKIGKGGKVKYKSLITLTKVGRNSIEISQOMKQPLQ---ANMPSISRG-----2227  
QY 1444 VRSRHTSVSSGSPVLRSTLHEAPKALSPCIYDDTSARRTPPVSYQNTMSRGSPPMNRTS 1503  
Db 2228 -----RTMIH-----IPGV-----RNSSTSPVSKKGP 2251  
QY 1504 DVTIPPNKSTNHERKSLITPT-QRESIPAK-SPVPYGVDPVYVSHSPFPDPIHRGSGTAGEVYV 1561  
Db 2252 PLKTPASKSPSEGTATTSRPAKPSVKSELSPVARQTSQIGGSSKAPSRSGSR-----2305  
QY 1562 SHLPTQLDPAMPFHRALDPAAMAYLFORQSLPTPGYPSQVQLYAMENTROTILNDYITSQ 1621  
Db 2306 -----DSTPSRAQOPL-----SR 2319  
QY 1622 QMOVNLRPDVA---RGLSPREPGLPLYPATRGIIIDLTNNPPTILVPHPGCTSTPPMDRI 1678  
Db 2320 PIQSPGRNISTPGRNGISLPNK-----LSQLPRT-SSPSTASTKSSGGSKM 2364  
QY 1679 TYIPGTQITPPRPYNASMSPGHPTHLAAAASARERERERE-----KERER 1726  
Db 2365 SY-----TSPGRQMSQONLT--KQTGLSRKNASSIPRSESASKGLNOMNNGANGANKVEL 2416  
QY 1727 ERIAAASDYLVRPGSQOPRGSHGVYRSPSVRTOETMLQORPSVFGQNTGTSVITP 1786  
Db 2417 SRMSSTKSS-----GSESD-----RGERPVLVRQSTFIKEAPS-----2449  
QY 1787 LDPTAQLRIMPLPAGGFSISQGLPASRYNTAADAALALVDAASAPOMDYKTESKHEA 1846  
Db 2450 --PTLR-RKLEESAFSELS---PSSRPASPTRSQAOPTVLSPLDPMLSL-THSSVQAG 2502

QY 1847 A--RLEENLRSAVSEQOOLEQKLTLEVEKRSVQCLYTSSAFPSGKQPQHSSVVYSEAG 1904  
Db 2503 GWRKLPPNL---STIEYNDGRPAKRHDIAI-----SHSESPRLINSGTWKREHS 2552  
QY 1905 KDKGPPKPSRYEELRTRGKTTITAAI-----FIDVIITRQIASDKDAREGSQSSD 1956  
Db 2553 KHSSSLP--RVSTWRRTGSSSILSASSESEKAKSEKDEKHVNSISGTKQSKENQVSAK 2610  
QY 1957 SSSLSLSHRYE-TPSDALEVISPASSPAPPOEKLOTQYQEVVYVQANQAEKNDTRQYEGPLH 2015  
Db 2611 TWRIKIKENEFPTNSTQTVSSGATNGAESKTLIYQMAPAVSK---TEDVWRIEDCDIN 2667  
QY 2016 HYPQOESPSPQOOLPP-----SSQAEGMQVPRTHRLITLADHI 2055  
Db 2668 NPR---SGRSPGTGTPVIDSVSEKANPNKIDSKDNQAKQNVGNSVPM--RTVGLNRL 2722  
QY 2056 CQITQDFARNQVSOPTPOQPT--STFONSALVST---PVRTKTSNRISPSQAOQSV 2110  
Db 2723 TSFIQVD-----APDQKGTETKPGQNNPVPVSETNESPIVERT-----PFSSSSS 2768  
QY 2111 HQORP---GSRVSPENLVDKSRGSRGCKSPERSHVSEPEYEPISPPQVYVH--EKODS 2164  
Db 2769 KHSSPSTGTVAAARVTPFN-----NPSPKSSADSTSARPSOIP-TPVNNNTKRRDS 2818

RESULT 5  
US-08-452-654-2  
; Sequence 2, Application US/08452654  
; Patent No. 5691454  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: KINZLYN, GEOFF  
; APPLICANT: KINZLER, KENNETH  
; APPLICANT: MARKHAM, ALEXANDER F.  
; APPLICANT: NAKAMURA, YUSUKE  
; APPLICANT: THLIVERIS, ANDREW  
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
; NUMBER OF SEQUENCES: 94  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Banner, Birch, McKie & Beckett  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001-4598  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/452,654  
; FILING DATE: 25-MAY-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/741,940  
; FILING DATE: 08-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 1107.035574  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2843 amino acids  
; TYPE: amino acid

|    |      |                                                               |      |
|----|------|---------------------------------------------------------------|------|
| Qy | 347  | PETRKOREQE-RFQVORGAGLSATIARSH-----EISEIIDLGSQENNEK            | 395  |
| Db | 981  | PSIESYEDDESFCYSQYPADLAHKHSANMHMDNDGELDTPINYSLKYSDEQLNSGR      | 1040 |
| Qy | 396  | QMRQLSVIPMMFMDAQRVKFINNGLMEDPMKYVKDRQFMWYTDHE---KEIFKDKFI     | 453  |
| Db | 1041 | Q-----SPSQNERWARPKHIEIDKIQSEQRQSRNOSTYVPVYTESTDDKHL           | 1087 |
| Qy | 454  | QHPKNFGLIASYLERKSDVCLYLYTKKNENYKALVRRNYKRRGRNQOIARPSOEBK      | 513  |
| Db | 1088 | KFQPHFG-----QQECVSPY-----RSRGANGSETNRVSGNHGINQVNSLQOE-        | 1132 |
| Qy | 514  | VEBKEEDKAEKTEK---EBEKKDEBEKDEKDEKENTKEKDKIDGTAEETEEREQATPRG   | 571  |
| Db | 1133 | -DDYEDDKPTNYSERYSEEEQHEEERPTNYSIKYN-EKRRHVQDQIDYSLKYATDIPSS   | 1190 |
| Qy | 572  | R-----KTANSQGRKRGITRSMTEAAAAAS-----AAAAAATEPPPPPPP            | 614  |
| Db | 1191 | OKOSFSSKSSSQSQSKTEBHMSSSENTSTPSSNAKRONQLHPSSAQRSGQPOKAATCK    | 1250 |
| Qy | 615  | PEPISSTEPVET-----SRWT-----EEMEVAKKGLVEHGRNNAALAKM---          | 653  |
| Db | 1251 | VSSINQETIQYCVEDTPICFSRCSSLSLSAEDEICGNOTTOEADSANTLQIAEIKGK     | 1310 |
| Qy | 654  | VGTKSE-----AQCKNPFYFYKRRHNDNLQOQHK-----OQTSR                  | 688  |
| Db | 1311 | IGTRSAEDVPYSEPAVVSQHPRTKSSRLQSSLSSESARKHAFEPFGSAKSPSKSGAQTPK  | 1370 |
| Qy | 689  | KPREE-----RDVSOCEVASTVSAQEDIEDIASNENEENDESEVAKPSE---DSPEN     | 740  |
| Db | 1371 | SPPEHYVQETPLMFSTCTSV-SLSOSFERSIASSVSEPCSGMVSGIISPSDLDPSPGQ    | 1429 |
| Qy | 741  | A--TSRGNTAPEVLEPFTTETAPSTSPSLAVPSTKPADESVEVETQVNDISIAETAEQMDV | 798  |
| Db | 1430 | TMPPSRSKTPPP---PPQTAQTKREVPKPKAKPTAEKRESGPKQAVNAVQ-RVQVLPDA   | 1485 |
| Qy | 799  | DOQEHSAAE-----GSVC-----DPPPAKQADSVDEVR-----VPEN-HASKVBDGN     | 840  |
| Db | 1486 | DTLLHFATESTPDGFCSSSLSALSIDEFFIQK-----DVELRIMPPVQENDNGNETESEQ  | 1541 |
| Qy | 841  | TKE--RDLDRASEKVEPRDELVAQAQINQAORPEQSDNDSSATCSADEVD-----       | 890  |
| Db | 1542 | PKGSNENQEKAEKTIIDSEKDLL-----DSD-----DDDIIELECIIIS             | 1581 |
| Qy | 891  | -----GEPERQRMF-----PMDSKPSLLNPTGSIIVSSLPKPNLDLPOLQHRAAVI      | 937  |
| Db | 1582 | AMPTKSRGKKPAQATASKLPPPVARKPSQL-PVYKLLPSQ-----NRL---QPOKHVSFT  | 1633 |
| Qy | 938  | P-----PMVSC-----TPONIPIGTPVSGYAL-----YQRHIKAMHE               | 969  |
| Db | 1634 | PGDDMPRVYCEGTPINFSTATSLSDLTIESPPNELAAGEVGRGAQSGEKEKRDTPTE     | 1693 |
| Qy | 970  | SAL-----LEBORQROBQIDLECRSSTSPGCTSKSPNREWEVL-----              | 1007 |
| Db | 1694 | GRSTDQAQKGTSSVTIPELDNDKNAEBGDIAEACINSAMPKGKSHKFRVKKIMQVQQA    | 1753 |
| Qy | 1008 | ---QAPAPHOLITNLPGVRLTPTRTPPPPLI-----                          | 1038 |
| Db | 1754 | SASSAPNK---NOLDGKKKKTSPVKPIPQWTEYTRVRKNADSKNNLNAERVFSNCKD     | 1810 |
| Qy | 1039 | SKKTTVASEKPSF-----IMGGSIS-----OQTPGTQYLTSHNQASYTOETP          | 1079 |
| Db | 1811 | SKKONLKNNSKFNDKLPNNEDVRGSAFDSPHHYTIECTP---YCFSRND-----        | 1860 |





Db 2181 IESESGIKGGKVKYKSLITGKVRNSSEISGMQKPLQ---ANMPSISRG----- 2227  
QY 1444 VRSRHTSVSSGSPVLRSTLHEAPKAQLSGIYDDTSARRTPVSYQNTMSRGSPMMRTS 1503  
Db 2228 -----RTMHH-----IPGV-----RNSSTSPVSKGP 2251  
QY 1504 DVTIPPNKSTNHERKSTLPT-QRESIPAK-SPYGVGDVSVSHSPDPHHRGSGAGEVYW 1561  
Db 2252 PLKTPASKSPSEGATTSRPGAKPSVKSELSPVARTSQIGGSKAPSRSGR----- 2305  
QY 1562 SHLTQDLPAMPFRHALDPAAAYLFORQLSPTPGYSQVQOLYAMENRTOTILNDYITSQ 1621  
Db 2306 -----DSTPSRAQOPL-----SR 2319  
QY 1622 QMOYNLRPDVA---RGLSPREOPLGLYPATRGIIIDITNMPPTILVPHPGCTSTPPMDRI 1678  
Db 2320 PIQSPGRNSISPGNGISPPNK-----LSQLPRT-SSPSTASTSSGSGKM 2364  
QY 1679 TYIPGTQITPPRPYNSASMSGHPTHLAAAASAEERERERE-----KERER 1726  
Db 2365 SY-----TSPGRMQSQNLTI-KQTGLSKNASSIPRESASAKGLNQMNNGANGANKVEL 2416  
QY 1727 ERIAAASDLVLRGSGOPRPGSHGVYSPSPSVRTQETMLQORPSVFOGTNGTSTVTP 1786  
Db 2417 SRMSSTKSS-----GSESD-----RGERPVLVRQSTFIKEAPS----- 2449  
QY 1787 LDPTAQLRIMPLPAGGPSISQGLPASRYNTAADALAAALVDAASAPOMDYKTESKHEA 1846  
Db 2450 --PFLR-RKLEESASFELS---PSSRPASPTRSQOATPVLSPDPMSLS-THSSVQAG 2502  
QY 1847 A--KLEENLRSAVAEBOOQLEKTEVKEKRSVQCIYTSAPPSGKPPHSSVYVSEAG 1904  
Db 2503 GWRKLPENL---SPTIENDGRPAKRHDIAI-----SHSESPSLPINRSGTWKREHS 2552  
QY 1905 KDKGPPPKSYEEELRGRKTTITAA-----FIDVITITROIASDKDARERGSOSSD 1956  
Db 2553 KHSSSLP--RVSTWRRGTSSSILSSESESEKAKSEKHSNISGTOKSKENQVSAGK 2610  
QY 1957 SSSSLSSHRYE-TPSDAIEVISPASSPAPQEKLTQYQEVVKANQANDPTROYEGPLH 2015  
Db 2611 TWRIKIKENESPTNSTQTVSSGATNAESKTLIYQMAPAVK---TEDVWVRIEDCPIN 2667  
QY 2016 HYRQOQSPSPQOOLPP-----SSAEGMGQVPRTHRLITLADHI 2055  
Db 2668 NPR---SGRSPTGTPVIDSVSEKANPNIKDKDNQAKQVNGSVPM--RTVGLENRL 2722  
QY 2056 CQIITQDFARNOVSSOTPOQPPT--STFQSPSALVT--PVRTKTSNRYSPESQAQSV 2110  
Db 2723 TSFIQVD-----APDQGTGIKQGNPVPVSETNESPIVERT-----PFSSSSS 2768  
QY 2111 HHQRP-----GSRVSPENLVDKSRGSRPGKSPERSHVSSEPEYSPQPVPVH--ERQDS 2164  
Db 2769 KHSSPSGTVAARVTPFY-----NPSRKSSADSTSNARPSQIP-TPVNNNTKKRDS 2818

## RESULT 7

US-07-741-940-7  
; Sequence 7, Application US/07741940  
; Patent No. 535275  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: JOSLYN, GEOFF  
; APPLICANT: KINZLER, KENNETH  
; APPLICANT: MARKHAM, ALEXANDER F.  
; APPLICANT: NAKAMURA, YUSUKE  
; APPLICANT: THLIVERIS, ANDREW  
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
; NUMBER OF SEQUENCES: 94

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner, Birch, McKie & Beckett  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001-4598  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/741,940  
; FILING DATE: 19920109  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 1107.035574  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2842 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: APC  
; US-07-741-940-7

Query Match 2.9%; Score 365; DB 1; Length 2842;  
Best Local Similarity 19.0%; Pred. No. 6.6e-14;

Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

QY 347 PETRKOREQE-RFQRVQGRGAGLSATIRSEH-----EISEIIDGLSEQENNEK 395  
Db 980 PSIESYSEDDSEFCSYQYPADLAHKHSANHMDDNGELDTPIYSLKYSDEQLNSGR 1039  
QY 396 QMQLSVIIPMMFDAEORRYKFINMGLMEDPMKYKDRQFMVWTDHE--KEIFKDKFI 453  
Db 1040 Q-----SFSQNERWARPKHIIIEDEIKQSEQRSRNQSTYYPVYTESTDDKHL 1086  
QY 454 QHPKNFGLIASYLERKSVDPCLVLYITKKENYKALVRNRYKRRGRNQOIARPSQEEK 513  
Db 1087 KFQPHFG-----QCECVSPY---RSRGANGSETNRVGSNHGICNQVSQLQCE- 1131  
QY 514 VEEKEEDKAEKTEKK--EEBKDBEEKDEKEDSKENTKEKDKIDGTABETEEREQATPRG 571  
Db 1132 -DDYEDDKPTNYSEYSEEEQHEEERPTNYSIKYN--EKRHRVDQPIDYSLKYATDIPSS 1189  
QY 572 R-----KTANSQGRKGRITRSMTNEAAS-----AAAAAATEEPPPLPPP 614  
Db 1190 OKOSFSEKSSSQSKTEHMSSESSENTSPSSNAKQNLHPSSAQSRSGQPKAATCK 1249  
QY 615 PEPISTEPVET-----SRWTEEMEYAKKGLVEHGRN-----WAAIAKIM 653  
Db 1250 VSSINQETIQTVCYEDTPICFSCSSLSLSAEDIGCNOTTQEADSANTLQIAEIKK 1309  
QY 654 VGTKSE-----AOCKNFYFNKRRNLNLOOHK-----OQTSR 688  
Db 1310 IGRSAEDPVSEYPAVPSQHPRTKSSRLQGLSSSESARHRAVEFSSGAKSPSKGATPK 1369  
QY 689 KPREE-----RDVSCVESVASTVSAQEDIEDIASNEENPEDEVEAVKPS--DSPEN 740  
Db 1370 SPPEHYVQETPLMFSRCTSV--SSLDSFESRSIASSVQSEPCSCMVSGIISFSLDPSFGQ 1428

QY 741 A--TSRGNTPEAVELEPTTETAPSTPSLAVPSTKPAEDSVETQVNDISISAETAQMDV 798  
DB 1429 TAPPKSKTTPP---PPQTAQTKREVKNKAPTAKREKSGPKQAANAAYO-RVQVLPDA 1484  
QY 799 DQOEHSAEE---GSVC-----DPPATKADSVDEVR-----VPEH-HASKVEGDN 840  
DB 1485 DTLHFATFESTPDGFSKSSLSALSLEDFIQK-----DVELRIMPVQOENGNGETESEQ 1540  
QY 841 TKE--RDLDRASEKVEPRDEDLVVAQINAQPERPQSDNDSSATCSADEVD----- 890  
DB 1541 PKESNENGEAEKXIDSEKDL-----DSD-----DDIEIEECIIS 1580  
QY 891 ---GEPERQRMFPMDSKSLNPTGSIILVSPKPNPLD-----LP-----QLQHRAAVIP 938  
DB 1581 AMPTKSSRKAKKPAQATKSLPPVAR-----KPSQLPVYKLLPSQNLQPKQHVSTP 1633  
QY 939 ---PMVSC---TPCNIPGTPVSYAL-----YORHIKAMHES 970  
DB 1634 GDMPRVYCVGPTINFSTATSLDLTIESPNEAAGEVGRGAQSGEFEKRDPTIPEG 1693  
QY 971 AL-----LEQORQEOIDLECRSSTSPCGTSKSPNREWVL----- 1007  
DB 1694 RSTDEAOGKTSSTVITPELDNKAEGDILAEICINSAMPKSKHKPFRVKKIMDOVQOAS 1753  
QY 1008 ---QAPAHOLITNLPVGLPRTTRPPPLI-----PS 1039  
DB 1754 ASSAPNK---NOLDGKKKPTSPVKPIQNTVEYRVRKNADSKNNLNAERVFSDNKS 1810  
QY 1040 SKTTVASEKPSF-----IMGSSIS-----OGTPGYLTSHNOASTYQETPK 1080  
DB 1811 KQNLKNNSKDFNDKLPNNEDRVGSAFSDPHHYTPIEGTP---YCFSRND----- 1859  
QY 1081 PSVGSISLGLPROQESAKSATLPYIKOEFRSPRSONSQPEGLLVRAQHEGVVGTAGAIQ 1140  
DB 1860 -SLSLDDEDDVLSRKAELRAKENKESEAKVTSITE--LTSNQOSA---NKTQAI 1913  
QY 1141 EGSITRGTPTSKISVESTPSLRGSIQTGTPALQGTPIEALVK-----GS 1186  
DB 1914 KOPINRGQPKPILQKS-----TFPOSSKDI PDGAATDEKLQNFALENTPVCFSHNS 1967  
QY 1187 ISRM-----PIEDSSPEKREBAASKGHVIEKSGHI----- 1219  
DB 1968 LSSLSIDIDQENNNKENEPKETEPDQSGEPKQASGAPKSPHFVEDTPVCFSRNSSL 2027  
QY 1220 -LSVDN-----IKNARECTSPRTAHEITSLKRSYVESGNIKOGMSRESPVSAPLE 1270  
DB 2028 SLSIDSEDDLQECISSAMPKKPSRLKGDNEKHSRPMGIIIGEDUTLQDKIQRPDS 2087  
QY 1271 GLICRALPRGSHDLKERTVLGSIQGTPTATESPEDGLKYPKQIKRESPPIRAFEG 1330  
DB 2088 E---HGLSPDSENFDMKAIQEGANSIVSSLHQAAAAAC-----LSRQASSDSSTLSLKS 2139  
QY 1331 AITGKPKVDGTTIKEMGRSTHEIPRODILTQESRKPEVVQSTRTPIEGISOGTPIKF 1390  
DB 2140 GISLGSPP-----HLTPDOEKPFNTKNGPRILK-----PGEKSTLETAKI 2180  
QY 1391 DNNS-GOSAIKHNVKSLITGP-----SKLSRGM-PPLEITVPENIKVVERGYEDVKAGETV 1444  
DB 2181 ESEKGIKGGKVKYKSLITGVKNSSEISGQMKOPLQ---ANMFSIRG----- 2226  
QY 1445 RSRTSVVSSGSPVLSRTHLEAPKAQLSPGIYDDTSARRTPVSYQNTMSRSGSPMNRSTD 1504  
DB 2227 -----RTMIH-----IPGV-----RNSSTSPVSKKGP 2251  
QY 1505 VTIPNKTSTNHERKSTLPT--QRESIPAK-SPVPGVDVPSHSPDPHHRGSTAGEVYWS 1562  
DB 2252 LKTPASKSPSEGGTATTSPRGAKPSKSELSPVARQTSIQIGSSKAPSRSGR----- 2304  
QY 1563 HLPTQLDPAMFPHRALDPAAYLQFQRLSPTPCYPSQYQLYAMENTROTILNDYITSQ 1622  
DB 2305 -----DSTPSRPAQOPL-----SRP 2319  
QY 1623 MQVNLRPDVA---RGLSPREQLGLPYPATRGIIDLTMNPPTILVPHPGGTSTPPMDRIT 1679

DB 2320 IQSPGRNISPCRNIGISPPNK-----LSQLPRT-SSPSTASTKSSSGKMS 2364  
QY 1680 YIPGTQITFPFPPPNASMSGPHTHLAAAAAERERERE-----KERERE 1727  
DB 2365 Y-----TSPGRQMSQOQLT--KOTGLSKNASSIPRSESASKGLNMNGNANKKVELS 2416  
QY 1728 RIAAASSDLYLRPQSEQPRGSHGYVRSPSVRTQMLQOORPSVFOGTNGTNGTSVITPL 1787  
DB 2417 RMSSTKSS-----GSESD-----RSEPLVLRQSTFFIKEAPS----- 2448  
QY 1788 DPTAQLRIMPLPAGGPSISOGLPASRYNTAADAALAYDAASAAPQMDVSKTKESKHEAA 1847  
DB 2449 -PTLR-RKLEESASFESLS---PSSRPASPTRSQATQTVLSPSLPDMSLT-THSSVQAGG 2502  
QY 1848 --RLEENLRSAVSEQQOQLEQKLEVEKRSVQCLYTSSAFFPSGKPOPHSVVYSEAGK 1905  
DB 2503 WRKLPPNL---SPTEYNDGRPAKRHDIAI-----SHSESPSLPINRSGTWKREHSK 2552  
QY 1906 DKGPPPKSRYEELRTRGKTTITAAI-----FIDVITRQIASDKDARERGSQSDS 1957  
DB 2553 HSSSLP--RVSTWRTGSSSILSASSESEKAKSEDEKHVNSISGTSKQENQVSAKGT 2610  
QY 1958 SSSLSHRYE-TPSDAIEVISPASPAPQPBKLOTYQEVVKAQNAENDPTRQVEGPLHH 2016  
DB 2611 WRKIKENEFSTNSTSQTVSSGATNGAESKTLIYQMAPAVSK---TEDVWVRIEDCPINN 2667  
QY 2017 YRQOESPSPOQLPP-----SSQAEQMGQVPRTHRLITLIADHIC 2056  
DB 2668 PR---SGRSPGTGNTPPVIDSYSEKANPNIKDKONQARQNGVSGVPM--RTVGLENRLN 2722  
QY 2057 QIITODFARNQVSSQTP--QOPPTSTFONSPSALYS--TPVRTKTSNRYSPESQAOQSVHHOR 2114  
DB 2723 SFIQVDAPDQKTEIKPCQNNPVPVSEINSESIYERTPFSSSSSKHSPSGTVA----- 2777  
QY 2115 PGSRVSPENLVKSRGSRPGKSPERSHVSSEPYEPISPPQVPPVH--EKQDS 2164  
DB 2778 --ARVTPNY-----NPSRKSADSTARSQIP-TPVNNNTKKRDS 2817

## RESULT 8

US-08-289-548A-7  
Sequence 7, Application US/08289548A  
Patent No. 5648212  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERTIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Allegretti, LTD  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/289,548A  
FILING DATE: 12-AUG-1994

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107,46943  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2842 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: APC  
US-08-289-548A-7

Query Match 2.98; Score 365; DB 1; Length 2842;  
Best Local Similarity 19.08; Pred. No. 6.6e-14;  
Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;  
Qy 347 PEIRKQREQOE-RFORVGORGAGLSATLARSEH-----EISEIIDGLSEQENNEK 395  
Db 980 PSIESYEDDESKCSYQYPADLAHKIHSANHMDDNGELDTINYSKYSDLEQNSGR 1039  
Qy 396 QMRQLSVTPMMFADAEORRVKFINNGLMEDPMKYKDRQFMNVWTDHE--KEIFKDKFI 453  
Db 1040 Q-----SPSQNERWARPKHIIIEIKQEQSRQNSQSTTVVYTESTDDKHL 1086  
Qy 454 QHPKNFGLIASYLERKSPVDCVLYLYLTCKKENYKALVRNRYKRGRCNQOIARPSOEK 513  
Db 1087 KFOPHG-----QQECVSP-----RSRGANGSTNRVGNHGNQNVQSLOCQ- 1131  
Qy 514 VEEKEEDKAERTEK--BEERKDEEKEDEKSKENTKEKDKIDGTABETEEREQATPRG 571  
Db 1132 -DDYEDDKPTNYSRYSEEQHEEERTNYSIKYN-EERKRVDPIDYSLKYATDIPSS 1189  
Qy 572 R-----KTANGQRRKGRITRSMNFAAAS-----AAAAATEEPPLPPPP 614  
Db 1190 OKQSFSSKSSGSSGSKTEHMSSESENTSTPSSNAKRONLHPSSAQRSGQPKAATCK 1249  
Qy 615 PEPTISTEVEN-----SRWTEEMEYAKGLVEGRN-----WAAIAKM 653  
Db 1250 VSSINQETIQTICVEDPTICFRSSSSLSAEDIEGCGNTTQPADSANTLQIAEIKK 1309  
Qy 654 VGTKSE-----AQCKNFYENYKRRHLDNLLOHK-----QKTSR 688  
Db 1310 IGTRSAEDPVSEPAVSQHPRTKSSRLOGSSLSSESARHKAVERFSSGAKSPKSGAQTPK 1369  
Qy 689 KPREE-----RDVQSCBASTVSAQDEIDEASNEENEDPEVEAVKPE--DSPEN 740  
Db 1370 SPPEHYVQETPLMFSSCTSV-SLSDSFESRSIASSVQPCSPGMVSGIISPSDLPDPSGQ 1428  
Qy 741 A--TSRGTPEAVELEPTETAPSTPSLAVPSKPAEDSEVEFQVNDSSIAETAQMDV 798  
Db 1429 TMPSPRSKTPPP---PPQTAQTKREVKNKAPTAEKRESGPKQAAVNAVQ-RVQVLPDA 1484  
Qy 799 DQQHSABE-----GVC-----DPPATKADSDVDVEVR-----YPEN-HASKVEGDN 840  
Db 1485 DTLHFATESPTDGFSCSSLSALSLEPFLQK----DVELRIMPVQENDNGNETESEQ 1540  
Qy 841 TKE--RDLDRASEKVEPRDEDLVVAQQTNAORPEFPQSDNDSSATCSADEVD----- 890  
Db 1541 PKSENEOEKAEKTIIDSEKDLL-----DDSD-----DDDIEILEECIIS 1580  
Qy 891 ---GEPEQRMPDMSKPSLNLPTGSIILVSSPLKPNPLD-----LP-----QLQHRAAVIP 938  
Db 1581 AMPTKSSRKAKKPAQTASKLPPPPVAR-----KPSQLPVYTKLLPSQNRLOPKHVSFT 1633

Qy 939 ----PMVSC---TPCNIPIGTVPVSGYAL-----YORHIKAMHES 970  
Db 1634 GDDMPRYVCEGTPIINFSTATISLDLIESPPNELAAGEVGRGAQGEFEKRTIITEG 1693  
Qy 971 AL-----LEEQRQEQOIDELECRSSSTSPGOTSKSPNREWEVL----- 1007  
Db 1694 RSTDEAOGGKTSSVTIPELDNDKAEEDGDLAECINSAMPKGSHPKPRVKKIMDQVOQAS 1753  
Qy 1008 --QAPHQLTNLPGEVRLPTTRTRPPPLI-----PS 1039  
Db 1754 ASSAPNK---NQLDGKKKPTSPVKPIQNTYRTRVRKNADSKNNLNAERVFSDNKDS 1810  
Qy 1040 SKTTVASEKPSF-----IMGSGIS-----QGTPTGYLTSHNOASVQTETPK 1080  
Db 1811 KKNLKNNSKDFNDKLPNNEDRVGRGFAFDSPHHYTIEGTP--YCFSRND----- 1859  
Qy 1081 PSVGSISLGLPRQOESAKSATLPYIKQEEFSPRSQNSOPEGLLVYRAQHEGVYVRGTAGAIQ 1140  
Db 1860 -SLSLDFDDDDVDLSREKAEELRKAKEKSEAKVTSHTE--LTSNQQA---NKTQAI 1913  
Qy 1141 EGSITRGTPTSKISVESITPSLGRSITQGTALPOTGIPTALVK-----GS 1186  
Db 1914 KQPIRGOPAPILQKS-----TFQSSKDIPIRGAAATDEKLNFAIENTPVCFSHNS 1967  
Qy 1187 ISRM-----PIEDSSPEKGRGEEAASKGHYIEGCKSGHI----- 1219  
Db 1968 LSSLSLDIDQNNKNEPIKETEPDQSGEPSPKQASGYAPKSFVEDTPVCFSRNSSL 2027  
Qy 1220 -LSDYN-----IKWAREGTRSPRTAHEISLKRSTVESVEGNKQKMSMRSPVAPLE 1270  
Db 2028 SLSLSDSEDDLQECISSAMPKKKPSRLKGDNEKHSRNMGGILGEDLTLLDKDIQRPDS 2087  
Qy 1271 GLICRALPRGSPHDLKERTVLGSIIMOGTPRATTFESFEDGLKYPKQIKRESPPIRAFEG 1330  
Db 2088 E---HGLSPDSENDWKAIOEGANSIVSSLHQAANAAC-----LSRQSSDSDSILSKS 2139  
Qy 1331 AITKGPYDGIITTIKEMGRSIEIPRODILTQESRKTPVEVQSTRPIEGSISOGTPIKF 1390  
Db 2140 GISLGSFP-----HLTPDQEEKPFTSNKGPRIK-----PGEKSTLETKKI 2180  
Qy 1391 DNNS-GSATAKHNVKSLITGP-----SKLSRGM-PPLEIVPENIKVGRKYEDYKAGETV 1444  
Db 2181 ESEKIGKGGKVKYKSLITIKVRSNSENISGMQKPLQ---ANMPSISRG----- 2226  
Qy 1445 RSRITSVSVSGSPVLRSLTHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRGSPPMNRSTD 1504  
Db 2227 -----RTMIH-----IPGV-----RNSSSSTSPVSKKGPP 2251  
Qy 1505 VTIPPNKSTNHERKSTLTPT-QRESIPAK-SPVPQVDPVWVSHSPDPHHRGSTAGEVYWS 1562  
Db 2252 LKTPASKSPSEGOATVTPRGAKPSVKSELSPVARQTSQIGGSSKAPSRSGR----- 2304  
Qy 1563 HLPQLDPAMPFHRALDPAANAAYLFQRLSPTPGYPSOYQIYAMENTRQTILNDYITSQQ 1622  
Db 2305 -----DSTPSRAQQL-----SRP 2319  
Qy 1623 MOYNLRPDVA---RGLSPREQLGLPYPAIRGIIIDLINMPTILVPHPGCTSTPMDRIT 1679  
Db 2320 IQSPGRNISPGRNISPPNNK-----LSQLPRT-SSPSTASTKSSGSRMS 2364  
Qy 1680 YIPGTQITTFPPRPNYSASMSGPHPTHLAAAASAERERERE-----KERERE 1727  
Db 2365 Y-----TSPGRQMSQNLIT--KQTGLSKNASSIPRESASKGLNMNGNGANKVELS 2416  
Qy 1728 RIAAASDLYLRPGSEOPGRPGSHGYVRSPSPSVRTQETMLQRPSPVFGTNGTSTVITPL 1787  
Db 2417 RMSSTKSS-----GSESD-----RSRPFVLRQSTFIKEAPS----- 2448  
Qy 1788 DPTAQLRIMPLPAGGPSISQGLPASRYNTAADALAAALVDAASAPOMDVSKTKESKHEAA 1847  
Db 2449 -PTLUR-KRLEESASFESLS---PSSRPASPTRSQATPVLSPLPDMSLS--THSSVQAG 2502



Qy 1040 SKTTVASEKPSF-----IMGSIS-----QGTPTGYLTSHNOASYTQETPK 1080  
Db 1811 KKONLKNNSKDFNDKLPNNEDVRGSPAFDHPHYTPIEGTP--YCFSRND-----1859  
Qy 1081 PSVGSISGLPQROESAKSAILPYIKQEEFSPRSONSOPEGLLVRAQHEGVVRRTAGAIQ 1140  
Db 1860 -SLSLDFDDDDVDLREKAKELRAKENKESEAKVTSHTS--LTSNQOSA---NKTQAI 1913  
Qy 1141 EGSITRGTPTSKISVESIPSLRGSTIGCTPALPOTGITEALVK-----GS 1186  
Db 1914 KOPINRGQPKPILQKS-----TFQSKDIPDRGAATDEKLNFAIENTPVCFSHNS 1967  
Qy 1187 ISRM-----PIEDSSPEKGREAAKSHVIEGKSGHI-----1219  
Db 1968 LSSLSIDIOENNNKENEPKETEPDSDGEPKQASGYAPKSFHVEDTPVCFSRNSLS 2027  
Qy 1220 -LSVDN-----IKNARETRSPRTAHEISLKRSEVSEGNKQKQMSRESVPVAPLE 1270  
Db 2028 SLSIDSEDDLQECISSAMPKKKPKSRLLKGNKHSRNMGGILGEDTLDLKDIQRPDS 2087  
Qy 1271 GLICRALPRGSPHSDLKERTVLGSIIMOGTTRATTESPEDGLKYPKQIKRESPIRAPEG 1330  
Db 2088 E---HGLSPDSEDFWKAIQGANISVSSLLHQAAC-----LSRQASSDSLSLS 2139  
Qy 1331 AITKGPYDGTITIKEMGRSIEIPRQDILAQESRKTPEVVQSTRPIIEGSIQGTPIKF 1390  
Db 2140 GISLSPF-----HLPTDQEKPTTSNKGPRILK-----PGEKSTLETKKI 2180  
Qy 1391 DNNS-GQSAIKHNKVLITGP-----SKLSRGM-PPEIVPENIKVVERKYEDYKAGETV 1444  
Db 2181 ESEKGIKGGKVKVYKSLTLGKVRNSSETSGOMKQPLQ---ANMPSISRG-----2226  
Qy 1445 RSRHTSVSSGSPVLSRLTLEAPRAQLSPGIYDDTSARRTPVSYQNTMSRCSPMNRTSD 1504  
Db 2227 -----RTMIH-----IPGV-----RNSSSSTSPVSKKGGP 2251  
Qy 1505 VTIPPNKSTNHERKSTLPTT-QRESIPAK-SPVPQVDVPSVSHSPDFHHRGSTAGEVYWS 1562  
Db 2252 LKTPASKSPSEGOYATTSFPRCAKPSVKSELSFVARQTSQIGSSKAPSRSGR-----2304  
Qy 1563 HLPQLDPAMPFHRALDPAALYLFQRLSPTGPYQYQILYAMENTRQTILNDYITSQQ 1622  
Db 2305 -----DSTPSRPAQOPL-----SRP 2319  
Qy 1623 MOVNLRPDVA--RGLSPREQPLPYPATRGIDLTNMPPTILVPHPGGTSTPPMDRIT 1679  
Db 2320 IQSPGRNISGRNGISPPNK-----LSOLPRT-SSPSTASTKSSGSGKMS 2364  
Qy 1680 YIPGTQITFPPRPNYSMSMCPHTHLAAAASAERERERE-----KERERE 1727  
Db 2365 Y-----TSPGRMSQOQLT--KOTGLSKNASTIPRESASKGLNOMNGNANKVELS 2416  
Qy 1728 RIAAASDLYLRPGEOPRGSGHYVRSPSPSVRTQETMLQQRPSVFQGTNGTSVITPL 1787  
Db 2417 RMSSTKS-----GSESD-----RSERPVLVROSTFKEAPS-----2448  
Qy 1788 DPTAQLRIMPLPAGGSPSTISQGLPASRYNTAALAAALVDAASAPQMDVSKTESKHEA 1847  
Db 2449 -PTLR-RKLESASFESUS---PSRRPASPRSQAPQTPVLSPLDMSLS-THSSVQAGG 2502  
Qy 1848 --RLEENLRSAVSEQQOQLEQTEKRSVQCLYTSSAFFSPGPKPQHSSVYVSEAGK 1905  
Db 2503 WRKLPPNL---SPTIEYNDGRPAKRHDIA-----SHSESPPRLPINRSGTWKREHSK 2552  
Qy 1906 DKGPPPKSRYEEELRTRGKTTITAN-----FIDVITRQIASDKDARERSQSDDS 1957  
Db 2553 HSSSLP--RVSTWRTGSSSILSASSESEKAKSEKHNVSISGTSKQENOVSAKGT 2610  
Qy 1958 SSSLSSSHRYE-TPSDATEVISPASSAPPQEKLTQYQEVVKAQNAENDPTQVEGPLHH 2016  
Db 2611 WRKIKENEFSTNTSQTVSSCATNGAESKTLIYQMAVASK---TEDVWVRIEDCPINN 2667

Qy 2017 YRQOQSPSPQOQLPP-----SSQAEQMGOVPRTHRLITLADHIC 2056  
Db 2668 PR---SGRSPGTNPPVIDSVSEKANPNIKSDKNQAKQNGVSGVPM--RTVGLNRLN 2722  
Qy 2057 QIITQDFARNQVSSQTP-QQPPTTFQNSPALVS-TPVTKTNRYSPEQASQSVHHOR 2114  
Db 2723 SFIOVDAPDQKGTIKPQNNPVPVSETNESSIVERTPFSSSSSKHSSPGTVA-----2777  
Qy 2115 PGRSVSPENLVKSRGSRPKSPERSHVSSEPIESPQVPPVH--EKQDS 2164  
Db 2778 --ARVTFNY-----NPSPRKSSADSTSARPSQIP-TPVNNNTKKRDS 2817

## RESULT 10

US-08-452-655B-2  
; Sequence 2, Application US/08452655B  
; Patent No. 5783666  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: JOSLYN, GEOFF  
; APPLICANT: KINZLER, KENNETH  
; APPLICANT: MARKHAM, ALEXANDER F.  
; APPLICANT: NAKAMURA, YUSUKE  
; APPLICANT: THLIVERIS, ANDREW  
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
; NUMBER OF SEQUENCES: 102  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001-4598  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/452,655B  
; FILING DATE: 25-MAY-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,548  
; FILING DATE: 12-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/741,940  
; FILING DATE: 08-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 1107,49964  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2843 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-452-655B-2

Query Match 2.9%; Score 365; DB 1; Length 2843;  
Best Local Similarity 19.0%; Pred. No. 6.6e-14;  
Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

QY 347 PEIKOREQOE-RFORVGORGAGLSATIARSEH-----EISEIIDLGLSEONNEK 395  
Db 981 PSIESYSEDDSKSCYQYADLAHKAHSANHMDNDGELDTPTINYSLKYSDDLNSGR 1040  
QY 336 QMRQLSVIPPMFMDAQRVVKFNNINGLMEDPMKYIKDQPMNVWTDHE--KEIFKDKFI 453  
Db 1041 Q-----SPSONERWAPKPHIIEDEIKQSEQRQNRQSTTPVVTTESTDDKHL 1087  
QY 454 QHPNGLIASYLERKSPVDCVLYLLTKKNENKALVRYNRKGRGRNQOIARPSQOEK 513  
Db 1088 KFOPHF-----QOECVSP-----RSGANGSTNRVGSNHGINQVQSLOQE- 1132  
QY 514 VEEKEEDKAETK--BEEKKDEEKDEKEDSKENTKEKIDGTABETEEREQATPRG 571  
Db 1133 -DDYEDDKPTNYSRYSEEQHEEERTNYSIKYN-EERHVDOPIDYSUKYATDIPSS 1190  
QY 572 R-----KTANSQGRKRGRITRSMNAAAAA-----AAAAATEEPPPPPPP 614  
Db 1191 QKQSFSSKSSGSSGSSKTEHMSSESENTSTPSSNAKRONQLHPSSAQRSGQPKAATCK 1250  
QY 615 PEPITSTEPVE-----SRWTEEMEYAKKGLVEHGRN-----WAAIAKM 653  
Db 1251 VSSINQETIQTICVEDPICSRRSSLSLSAEDEICGNOTTOEADSANTLQIAETEK 1310  
QY 654 VGTKSE-----AQCKNFYNYKRRHNLNLLQOHK-----QKTSR 688  
Db 1311 IGTRSAEDPVSEPAVSQHPRTKSRLOGSSLSSESARHKAVERSSGAKSPKSGAQTPK 1370  
QY 689 KPRE-----RDVQSCEVASTVSAQDEEDIEASNEEENEPEDSEVAVKPSE--DSPEN 740  
Db 1371 SPPEHYVOETPLMFSRCTSV--SSLSDESRSIASVSQPCSGMVGIISPSDLPSDQ 1429  
QY 741 A--TSRGNTEPAVELEPTETAPSTPSLAVPSTKPADESEVETQVNDISAEATAEQMDV 798  
Db 1430 TMPSPRSKTTPP--PPQTAQTKREVFNKAPTAKEKRESGKGAANVAQV-RVOVLPA 1485  
QY 799 DOQHSABE-----GSVC-----DPPATKADSVDEVEYR-----YPEN-HASKVEGDN 840  
Db 1486 DTLHFATESTPDGFSKSSLSALSLEBPFIOK-----DVELRIMPPVQVONGNETESEQ 1541  
QY 841 TKE--RDLDRASEKVEPDEDLVAQOINQAORPEPQSDNDSATCSADEVD-----890  
Db 1542 PKSENQEKAEKTIIDSEKDLL-----DDSD-----DODTEILEECIIS 1581  
QY 891 ---GEPERQRMFMDKPSLLNPTGSLVSSPLKPNLD---LP-----OLQHRAAVIP 938  
Db 1582 AMPKSSRKAKKPAQTASKLPPVAR-----KPSQLPVYKLLPSQNRLOPQKHVSETP 1634  
QY 939 ---PMVSC-----TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
Db 1635 GDDMPRVYCVGTPINFSTATSLSDLTIESPPNELAAGEVGRGAQSGEPEKRDITIPEG 1694  
QY 971 AL-----LEQRORQEQIDLECRSSTSPCGTSKSPNREWEVL-----1007  
Db 1695 RSTDEAOGKTSVTIPELDNKAEBGDILAEICNSAMPKKGSHKPFVKKIMQVQOAS 1754  
QY 1008 --QAPHOLITNPEGRVLRTPTRTPPPPLI-----PS 1039  
Db 1755 ASSAPNK---NOLDGKKKFTSPVKPIQNTYETRVKRNKADSKNNLNAREVFSNDKDS 1811  
QY 1040 SKTTVASEKPSF-----IMGGSIS-----QGTPTGYLTSHNOASYTQETPK 1080  
Db 1812 KQNLKNNSKDFNDKLPNNEDRVRGSAFDSPHHYTPIEGTP--YCFSRND-----1860  
QY 1081 PSVGSISIGLPROQESAKSATLPYKQEEFSPRQNSQOPELLVRAQHEGVVRGTAGAIQ 1140  
Db 1861 -SLSLDFDODDDVLSREKAEFLKAKENKESEAKVTSHTE--LTSNQA-----NKTQAI 1914  
QY 1141 EGSITRGTPTS KISVESIPSLRGSITGTPTALPOTGIPTEALVK-----GS 1186  
Db 1915 KQPINRGQPRPILOQS-----TFPOSSKDIPIRGAAATDEKQNFALIENTPVCFSHNS 1968  
QY 1187 ISRM-----PIEDSSPEKGREEAASKGHVIEGKSGHI-----1219

Db 1969 LSSLSIDDOENNNKENEFIKETEPDSCQEFKSKPOASGYAPKSFHVEDTPVCFSRNSSLS 2028  
QY 1220 -LSYDN-----IKNAREGTRSPRTAHEISLKSRYSEVEGNKQGMRESVPYAPLE 1270  
Db 2029 SLSIDSEDDLQECITSSAPKKKKPSRLKGDNEKHSRNMGGIILGEDLTDLKDIORPDS 2088  
QY 1271 GLICRAIPRGSPHSDLKERTVLSGIMOGTTPRATESPEDGLKYPKQIKRSPPIRAPEG 1330  
Db 2089 E---HGLSPDSENFDMKAIQEGANSIVSSLHAAAAAC-----LSRQASDSDSLKS 2140  
QY 1331 AITKGPKGDTITTIKEMGRSHEIPRODILQESRKTPEVQVSTRPIIEGSTSQGTPK 1390  
Db 2141 GISLSP-----HLTPDOEEKPFTSNKGPRILK-----PGEKSTILETKI 2181  
QY 1391 DNNS-GOGAIIKHNKSLITGP-----SKLSRGM-PLEITVPENIKVVERGYEDVKAQETV 1444  
Db 2182 ESEKGIKGGKVKVYKSLITGKVRNSSETSGMKOPLO---ANMPSISRG-----2227  
QY 1445 RSRHTSVVSSGSPSVLRSLRHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRSGPMNRTSD 1504  
Db 2228 -----RTMIH-----IPGV-----RNSSTSPVSKKGPP 2252  
QY 1505 VTIPPNKSTNHERKSTLTPT-QRESIPAK-SPVPCVDVPVWSHSPDPHHRGSTAGEVYWS 1562  
Db 2253 LKTASKSPSEGTATTSRPGAKPSVKSELSFVARQTSQIGSSKAPSRSGSR-----2305  
QY 1563 HLPTOLDPMPFHRALDPAAYLFORLSTPGYPSQYQIYAMENTROTILNDYITSOQ 1622  
Db 2306 -----DSTSRPAQOPL-----SRP 2320  
QY 1623 MOVNLRPDVA--RGLSPREQPLGYPATRGIIIDLTMNPPTILVPHPGTSTPMDRIT 1679  
Db 2321 IQSGRNSISGRNGISPPNK-----LSQLPT-SSPSTASTKSSGGKMS 2365  
QY 1680 YIPQTOITFPPRPNYSASMPGHPHTLAAAAASAERERERE-----KERERE 1727  
Db 2366 Y-----TSPGRQMSQOHLT--KOTGLSKNASSIPRSESASKGLNMNGNANKKVELS 2417  
QY 1728 RIAAASSDLYLRPSEQPGRGSHGYRSPSPSVRTQETMLQORPSVFOGTNGTSVITPL 1787  
Db 2418 RMSSTKS-----GSESD-----RSERPVLRVROSTFKEAPS-----2449  
QY 1788 DPTAQLRIMPLPAGGPSISQGLPASRYNTAADAALADAAAPOMDVSTKESKHEAA 1847  
Db 2450 -PTLR-RKLEESASFESLS---PSRRPASPTRSQAQTPVLSPSLPDMXLS-THSSVQAGG 2503  
QY 1848 --RLEENLRSSAAVSEOOQLEQKTLVEKRSVQCLYTSSAPPSGKPOPHSSVYVSEAGK 1905  
Db 2504 WRKLPPNL---SPTIEYNDGRPAKRHDIAI-----SHSESPSRLPINRSGTWKREHKS 2553  
QY 1906 DKGPPPKSRYEELLRGKTTITTAAN-----FIDVITRQASDKDARERSQSSDS 1957  
Db 2554 HSSSLP--RVSTWRTGSSSILSASESESEKAKSEDEKHVNSISGTSKQENQVSAKGT 2611  
QY 1958 SSSLSSSHRYE--TPSDATEVISPASSAPPQEKLOTQOYEVVKAQNAENDPTROVEGLPH 2016  
Db 2612 WRKIKENEFTNTSQTYSVGATNGAESKTLIYQMAPAVSK---TEDVWTRIEDCPINN 2668  
QY 2017 YRPOQESPSPOOQLPP-----SSQAEGMGQVPRTHRLITLADHIC 2056  
Db 2669 PR---SGRSPGTNTPPVIDSVSEKANPNIKDSKNOAKQNVGNGSVPM--RTVGLENRLN 2723  
QY 2057 QIITQDFARNQVSSQTP-QQPTSTTFQNSPSALYS-TPVRTKTNRYSPESQAQSVHHQR 2114  
Db 2724 SFIQVADAPDQKGTIKPQNNPVPVSETNESSIVERTPFSSSSSKSHSSPGTVA-----2778  
QY 2115 PGSRVSPENLVKSGRSGKSPERSHVSSEPEPISPPQVPPVH--EKQDS 2164  
Db 2779 --AVTTPFNY-----NPSPKSSADSTARSQIP-TPVNNNTYKKRDS 2818

RESULT 11

US-08-452-655B-7

Sequence 7, Application US/08452655B  
Patent No. 5783666  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,655B  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/289,548  
FILING DATE: 12-AUG-1994  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.49964  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
US-08-452-655B-7

Query Match 2.98; Score 365; DB 1; Length 2843;  
Best Local Similarity 19.08; Pred. No. 6.6e-14;  
Matches 409; Conservative 289; Mismatches 807; Indels 648; Gaps 94;  
Qy 347 PEIRKQEQQR-RFQRVQGRGAGLSATIRSEH-----BISETIDGLSPQENNEK 395  
Db 981 PSIESSEDDSKSCYQYQADLAHKSANHMDNDGELDTPIYSLKYSQDLNSGR 1040  
Qy 396 QMRQLSVIPPMFAEORRVFINNGLMEDPMKYKDRQPMNVWTDHE--KEIFKQKFI 453  
Db 1041 Q-----SPSQNERWARPKHIEDEIKQSEQRQSRNQSTTPYVYTESTDDKHL 1087  
Qy 454 QHPKNFGLIASLYERKSPVDCVLYLTKKNYKALVRRNYGRRGRNQOIARPSOEK 513  
Db 1088 KQPHFG-----QQECVSPY-----RSRGANGSETNRVGNHGINQNVSLCOE- 1132

Qy 514 VEEKEDKAETK--EKKDEEKDEKSDKENTKEKIDCTAETREERQATPRG 571  
Db 1133 -DYEDDKPTNYSERYSEEEHEEERTNYSIKYN-EERHVDQPIDYSLKYATDIPSS 1190  
Qy 572 R-----KTANSQRRKGRITRSMTNEAAAAS-----AAAAATEEPPLPPPP 614  
Db 1191 QKQSFSEKSSGSSGSSKTEHSSSSSENTSTPSSNAKRNQLHPSSAQSRGQPKAAATCK 1250  
Qy 615 PEPISSTEPVET-----SRWTEEMEYAKKGLVEHGRN-----WAAATKM 653  
Db 1251 VSSINQETIOTIYCVEDTPICFSRCSLSLSAEDIEGCTQTQADSAANTLQIAETKEK 1310  
Qy 654 VGTKSE-----AOCKNFYNYKRRHNLNLOQHK-----OQTSR 688  
Db 1311 IGTRSAEDPVSEPAVSOHQHPTKSRLOGSSLSESARHKAVERFSSGAKSPSKSGATPK 1370  
Qy 689 KPREE-----RDVSOCSVASTVSAQEDIEDIASNEEENPEDSEVEAVKPSSE--DSPEN 740  
Db 1371 SPPEHYVOETPLMFSRCTSV--SSLDSEFSRSIASSVQSEPCSGMVSGIISPSDLPSPGQ 1429  
Qy 741 A--TSRGNTPEPAVELEPTTETAPSTPSLAVPSTKPADESVETQVNDSTSAETAQMDV 798  
Db 1430 TMPPSRSKTPPP-----PPOTAQTKREVPKNKAPTAEKRESGKQAAVNAVO--RVQVLPDA 1485  
Qy 799 DQOEHSAEE-----GSVC-----DPPPATKADSVDEVR-----VPEN-HASKVEGDN 840  
Db 1486 DTLHFATSTPDGFCSSSSLSALSDPEFTQK---DVELRIMPVQENDNGHETSEQ 1541  
Qy 841 TKE--RDLDRASEKVEPRDEDLVVAQOINAOPEPQSDNDSSATCSADEVD-----890  
Db 1542 PKESNEQKEAKTIDSEKDL-----DDSD-----DDOIELEECIIS 1581  
Qy 891 ---GEFERQRMFMDSKPSSLNPTGSLVSSPLKPNPLD---LP-----QLQHRAAVIP 938  
Db 1582 AMPTKSRKAKKPAQATASKLPPPVAR-----KPSQLPVYKLLPSQNRQPQKHVSFT 1634  
Qy 939 ---PMVSC---TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
Db 1635 GDDMPRYCVGEGTPIFSTATSLDITIESPPNELAAGEVGRGAQGEFEKRTITTEG 1694  
Qy 971 AL-----LEEQROQEOIDLERSSTSPGCTSKSPNREVL-----1007  
Db 1695 RSTDEAOGGKTSVTIPELDDNKAEEDGILAEICINSAMPKSHKPFVKKIMDOVQOAS 1754  
Qy 1008 --QPAPHOLITNLPEGVRLPTTRTPPPPLI-----PS 1039  
Db 1755 ASSAPNK---NQLBGKKKPTSPVKPIQONTETRYRVRKNADSKNNLNARVFSDNKDS 1811  
Qy 1040 SKTTVASEKPSF-----IMGSGIS-----QGTPTGTYLTSHNOASYTOETPK 1080  
Db 1812 KKQNLKNSKDFNDKLPNNEDRVGRSFAFDSPHHYTPIEGTP--YCFSRND-----1860  
Qy 1081 PSVGSISLGLPROQESAKSATLPIYKQEEFSPRSONSOPEGLIVRAQHEGVVRGTAGAIQ 1140  
Db 1861 -SLSLDFDDDDVDLSREKAEELRAKENKESEAKVTSHTE--LTSNQOSA---NKTOAIA 1914  
Qy 1141 EGSITRGTPTSKISVESIPSLRGSITOGTPALPOTGTPTEALVK-----GS 1186  
Db 1915 KQPINRQOPILQKS-----TFPOSSKDIPDGAATDEKLNFAIENTPVCFSHNS 1968  
Qy 1187 ISRM-----PIEDSSPEKGREAAASKGHVYIEGKSGHI-----1219  
Db 1969 LSSLSLDIDQENNNKENEPKETETPPDSQGEPSKPAQSGYAPKSPHVEDTPVCFSRNSLS 2028  
Qy 1220 -LSVDN-----IKNAREGTRSPRTAHSLSKRSVESVEGNKQGMRESVPSAPLE 1270  
Db 2029 SLSIDSEDDLQECISSAMPKKKPSRLKGDNEKHSRNMGGIILGEDLTLDLKDQIAPDS 2088  
Qy 1271 GLICRALPRGSHDLKERTVLSGSIQMGTPRATTFESFEDGLKVPKQIKRESPPIRAFEG 1330  
Db 2089 E---HGLSPSENFEDWAKIQEAGANSIYSSLHQAAAAC-----LSRQASSDSLSILSKS 2140



Qy 1331 AITKGPYDITIKEMGRSIEIPRODILQESRKTEPVVQSTRPIIEGSIQCTPIKF 1390  
Db 2141 GISLGSPE-----HLTPDQEKFPSTKNGPRILK-----PGEKSTLETKKI 2181  
Qy 1391 DNNS-GQSAIKHNKSLITGP-----SKLSRGM-PPLIEIVPNIKVVERGKEDYKAGETV 1444  
Db 2182 ESEKGIKGGKVKYKSLTIGVRNSENSEISGOMKQPLQ---ANMPSISRG-----2227  
Qy 1445 RSRUTSVVSSGSPVLSRSTLHEAPKAQLSPGIYDDTSARRTPVSVQNTMRSGSPMMNRISD 1504  
Db 2228 -----RTWIIH-----IPGV-----RNSSTSPVSKGKPP 2252  
Qy 1505 VTIIPNKSTNHERKSTLTPT-QRESIPAK-SPVPGVDVWVSHSPFDPHHRGSTAGEVYWS 1562  
Db 2253 LKTPASKSPSEGQTATTSRCAKSPVSKSELSVPAQTQSIGGSSKAPSRSGR-----2305  
Qy 1563 HLPQLDPAMPFHRLDPAALPAAAYLFQRLSPTPGYPSQYQILYMENTQTILNDYITISQ 1622  
Db 2306 -----DSTPSRPAQOPL-----SRP 2320  
Qy 1623 MOVNLRPDVA---RGLSPREOPLGLPYPATRGIIDLTNMPPTILVPHPGGTSTPMDRIT 1679  
Db 2321 IQSPGRNISGRNGISPPNK-----LSQLPT-SSPSTASTKSSGSGRHS 2365  
Qy 1680 YIPGTQITFPPRPYNASMSGPHPTLAAASAERERERE-----KERERE 1727  
Db 2366 Y-----TSPGRMSQOQLT---KOTGLSKNASSIPRSSESASGLNOMNNGANKKVELS 2417  
Qy 1728 RIAAASSDLYLRPGSEQPRGSHGYRSPSPSVRTQETMLQORPSVFOGTNGTSVITPL 1787  
Db 2418 RMSSTKSS-----GSESD-----RSERPVLVROSTTFKEAPS-----2449  
Qy 1788 DPTAQLRIMPLPAGPSPISQGLPASRYNTAADAALVDAASAPQMDVSKTKESKHEAA 1847  
Db 2450 -PTLR-RKLEESASFELS---PSSRPASPTRSQATPVLSPLDMSLS-THSSVQAGG 2503  
Qy 1848 --RLEENLRSAVSEOOQLEQKTELEKRSVOCCLYTSSAFFPSGKPOPHSVVYSEAGK 1905  
Db 2504 WRKLPPNL---SPATIEYNDGRPAKRHDIAI-----SHSESPSRLPTNRSCTWKREHSK 2553  
Qy 1906 DKGPPPKSRYEEELTRCKTTITAN-----FIDVITRTASDKDARERGSSDS 1957  
Db 2554 HSSSLP--RVSTWRTGSSSILGASSESEKAKSEDEKHVNSISGTKSKENQVSAKGT 2611  
Qy 1958 SSSLSHRYE-TPSDAIEVISPASPAPPQEKLTQYQEVVKAQNAENDPTRQYEGPLHH 2016  
Db 2612 WRKIKENEFSTNSTQTSVSGAINGAESKTLIYQMAPAVSK---TEDVWVRIEDCPINN 2668  
Qy 2017 YRPOQSPSPQOQLPP-----SSQAEGMGQVPRTHRLITILADHIC 2056  
Db 2669 PR---SGRSPGTNTPPVIDSVSEKANPNIKDSKONQAKQNVGNSVPM--RTVGLNRLN 2723  
Qy 2057 QIITQDFARNQVSSQTP-QQPTSTFQNSPALYS-TPVTKTNNRYSPEQAOVSHOR 2114  
Db 2724 SFIOVDAPDQKGTETKPGQNNPVPVSETNESSIVERTPFSSSSSKHSSPSGTVA-----2778  
Qy 2115 PGSRVSPNLVDKSRGSRGKSPKSHVSSPEYEPISPPQVPPVH--EKQDS 2164  
Db 2779 --ARVTPFNY-----NPSPKSSADSTISARPSQIP-TPVANNTKKRDS 2818

## RESULT 12

US-08-450-582-2

; Sequence 2, Application US/08450582

; Patent No. 6114124

; GENERAL INFORMATION:

; APPLICANT: ALBERTSEN, HANS

; APPLICANT: ANAND, RAKESH

; APPLICANT: CARLSON, MARY

; APPLICANT: GRODEN, JOANNA

; APPLICANT: HEDGE, PHILIP J.

; APPLICANT: JOSLYN, GEOFF

; APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERTS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,582  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/452,655  
FILING DATE: 25-MAY-1995  
APPLICATION NUMBER: US 08/289,548  
FILING DATE: 12-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.49964  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-450-582-2

Query Match 2.9%; Score 365; DB 3; Length 2843;

Best Local Similarity 19.0%; Pred. No. 6.6e-14;

Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

Qy 347 PEIRKQEQQE-RFQVVGQRGAGLSATIRSEH-----EISEIDGLSEQENNEK 395  
Db 981 PSIESYDEDESKFCSYQYPADLAHKIHSANHMDNDGELDTPINYSIKYSDQELNSGR 1040  
Qy 396 QMRQLSVIPPMFADAEQRRVKFINNGLMEDPMKYKDRQFMNVTDHE--KEIFKDKFI 453  
Db 1041 Q-----SPSQNERWARPKHIEDEIKQSEQRQSRNQSTTYPVYTESTDDKHL 1087  
Qy 454 QHPKNFGLIASYLERKSVDPDCLVLYLTKKNENTKALVRRNYKRRGRNQIARPSQEEK 513  
Db 1088 KFQPHFG-----QQECVSPY-----RSRGANGSETNRVGNHGINQVSLCQE- 1132  
Qy 514 VEEKEEDKAEKTEKK--EERKKDEEKEDEKSDENTKEKDKIDGTAETETEREQATPRG 571  
Db 1133 -DDYEDDQPTWYSERYSEEEQHEEERTNYSIKYN-BEKRHVDDQPIDYSUKYATDIPSS 1190  
Qy 572 R-----KTANSQGRKGRITRSMTEAAAAA-----AAAAAATPEPPPLPPP 614  
Db 1191 QKQSFSEKSSSGOSSKTEHMSSESSENTSTPSSNAKRONQLHPSSAQSRSGQPQKAATCK 1250  
Qy 615 PEPITSTEPVET-----SRWTEEMEVAKKGLVEHGRN-----NAAIAKM 653  
Db 1251 VSSINQETIQYCVEDTPICFSRCSLSLSSAEIDGICNQTQEQADSANTLQIAEIKEK 1310

Qy 654 VGTKE-----AOCKNFYFKRRHNDLLOQHK-----QKTSR 688  
Db 1311 IGTRAEEDPVSEVPAVSQHPRTKSRLOGLSSLSBSARHKAVERSSGAKSPSKSGAOTPK 1370  
Qy 689 KPREE-----RDVSCQESVASTYSAQDEDEIASNEBENEDSEVEAVKPSE--DSPEN 740  
Db 1371 SPPEHYQVETPLMFSRCTSV--SSLDSPFSRSTASSVQSEPCSGVMGVIISPSDLDPSPGQ 1429  
Qy 741 A--TSRGMTEPAVELEPTETAPSTPSLAVPSTKPADESVETQVNDISIAETAQMDV 798  
Db 1430 TMPFSRSKTPPP--PPQTAOTKREVKNKAPTAEKRESGPKQAAVNAVQ--RVQVLPDA 1485  
Qy 799 DQOEHSAEE--GSMC-----DPPATKADSVDEVR--VPEN--HASKVEGDN 840  
Db 1486 DTLHFATETSPDGFSCSSLSALSDEFFIQ--DVELRIMPVOENDONGNETESEQ 1541  
Qy 841 TKE--RDLDRASEKVEPDEDLVVAQINAQRPQSDNDSSATCSADEVD-----890  
Db 1542 PKESNEQKEAETIDSEKOLL-----DDSD-----DDDEILEECIIS 1581  
Qy 891 --GEPEQRMPMDSKSLNPTGSIIVSSPLKPNPLD--LP-----QLQHRAAVIP 938  
Db 1582 AMPTKSSKAKKPAQATSKLPPVAR-----KPSQLPVYKLLPQNRLQPKQHVSTP 1634  
Qy 939 --PMVSC--TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
Db 1635 GDDMPVYCVGCTPINFSTATSLDLTIESPPNELAAGEVGRGAQSGEFKRTIPTTEG 1694  
Qy 971 AL-----LEBORQOEIDLECRSSTSPCGTSKSPNREWL-----1007  
Db 1695 RSTDEAOGKTSSTVPELDDNKAEEGDILAEICINSAMPKSKHKPRVKKIMQVQOAS 1754  
Qy 1008 --QAPHOLIINLEGVRLPTRTRPPPLI-----PS 1039  
Db 1755 ASSAPNK--NOLDGKKKTPSVKPIQNTYRTRVRKNADSKNNLNAERVFSDNKDS 1811  
Qy 1040 SKTVASEKPSF-----IMGGSIS-----QGTPTGYLTSHNQASYTQETPK 1080  
Db 1812 KKONLKNNSKDFNDKLPNNEDVRGSAFADSPHHYTPTEGTP--YCFSRND-----1860  
Qy 1081 PSVGSISGLPQOESAKSATLPYKQEEFSPRSONSOPEGLLVRAQHEGVVRRTAGAIQ 1140  
Db 1861 --SLSLDDDDVDLSREKAELRAKENKESEAKVTSHT--LTSNQOSA--NKTOAIA 1914  
Qy 1141 EGSITRGPTSKISVESTPSLRGSIQTCTPALPOTGIPTEALVK-----GS 1186  
Db 1915 KQPINRGQPKILQKS-----TFQSSKDI PDGGAATDEKLNFAIENTPVCFESHNS 1968  
Qy 1187 ISRM-----PIEDSSPEKGREAAKSHVYIEGKSGHI-----1219  
Db 1969 LSSLSDIDQENNNKENEPKETEPPDSQGEPSKPOASGYAPKSPHVEDTPVCFSRNSSL 2028  
Qy 1220 --LSYDN-----IKNARETRPRTAHEISLKRSEVSEGNKQGMRESVPSPAPLE 1270  
Db 2029 SLSIDSEDLLOECISSAMKKKPSRLKGNKHSRNMGIILGEDITLDLKDIOQRPDS 2088  
Qy 1271 GLICRALPRGSPHSLKERTVLGSIIMGTPTATSEFEDGLKVPKQIKRESPPITAFEG 1330  
Db 2089 E--HGLSPDSENFDAKIQAGANSIVSSLHQAAAAC-----LSRQASSDSLSLS 2140  
Qy 1331 AITGKPKYDGIITIKEMGRSITHEIPRODILQOESKRKTPVVQSTRPIEGSISQGTPIK 1390  
Db 2141 GISLGSPP-----HLTPDOEKPKFTSNKGPRLK-----PGEKSTLETKKI 2181  
Qy 1391 DNNS--GQSAIKHNKSLITGP-----SKLSRGM--PPELVIPENIKVVERGKYEDYKAGETV 1444  
Db 2182 ESEKGIKGGKVKVYKLLITKVRNSENSEISQMKQPLQ--ANMPSISRG-----2227  
Qy 1445 RSRHSTSVVSSGPSVLRSTLHEAPKAPLSPIGYDDTSARRTPVSYQNTMGRSGPMNRITSD 1504  
Db 2228 -----RTWIH-----IPGV-----RNSSTSPVSKKGP 2252

Qy 1505 VTIPPNKSTNHERKSTLTPT--QRESIPAK--SPVPGVDVPSHSPFDPHHRGSTAGEVYS 1562  
Db 2253 LKTPASKPSBEGQATATSPRGAKFSVKSELSFVARQTSQIGGSKAFSRSGSR-----2305  
Qy 1563 HLPTQLDPAMPFHRALDPAAAYLFQRQLSPTPGYPSQYQLYAMENTROTILNDYITISQ 1622  
Db 2306 -----DSTPSRPAQOPL-----SRP 2320  
Qy 1623 MQVNLRPDVA---RGLSPREQPLGLPYPATRGIIDLNMPTILVPHPGGTSTPMDRIT 1679  
Db 2321 IQSPGRNISIPGRNGISPPNK-----LSQLPRT--SPSTASTKSSGSGKMS 2365  
Qy 1680 YIPGTQITFPPRPYNSASMSGHPHTLAAAASAEERERERE-----KERERE 1727  
Db 2366 Y-----TSPGRMSQOQLT--KOTGLSKNASSIPRESASKGLNOMNGNCAKKVELS 2417  
Qy 1728 RIAAASDLYLRPGSEQGRPGSHGYVRSPSPSVRTQETMLQQRPSVFQGTNGTSTVITPL 1787  
Db 2418 RMSSTKSS-----GSED-----RSERPVLVRQSTFKEAPS-----2449  
Qy 1788 DPTAQLRIMPLPAGGPSISQGLPASRYNTAADALAALVDAASAPQMDVSKTKESKHEAA 1847  
Db 2450 --PTLR--RKLEESASFESLS-----PSSRPASPTRSQOATPVLSPLDMSLS--THSSVQAGG 2503  
Qy 1848 --RLEENLRSSAAVSEQQLEQKTLVEKRSVQCLYTSSAFFSGKPKQPHSHSVVYSEAGK 1905  
Db 2504 WRKLPPNL--SPTEYNDGPPAKRHOIAR-----SHSESPRLPINRSGTWKREHSK 2553  
Qy 1906 DKGPPPKSRVEEELRTGKTITTAAN-----FIDVITQIASDKDAREGSSQSDS 1957  
Db 2554 HSSSLP--RVSTWRTGSSSILSASSESEKAKSEKHNVSISGTSKQSKENOVSAKGT 2611  
Qy 1958 SSSLSHRYE--TPSDATEVISPASSPAPPQEKLOTYQEVVKANQAEENDPTROYEGPLHH 2016  
Db 2612 WRKIKENEFSPTNOSTVSSGATGAEKTLIYOMAPAVSK---TEVWVRIEDCPINN 2668  
Qy 2017 YRQOESPSPOQQLPP-----SSQAEQMGVQVPRTHRLITLADHIC 2056  
Db 2669 PR---SGRSPGTNTPVIDSVSEKANPNIKSDKNQAKQNVGNSVPM--RTVGLENRLN 2723  
Qy 2057 QITQDFARQNVSSQTP--QQPTTFQNSPALSYS--TPVTKTSNRYSPEQASQSVHHQR 2114  
Db 2724 SFIOVDAPDQKGTIKPQNNPVVSETNESSIVERTPFSSSSSKSHSSPSGTVA-----2778  
Qy 2115 PGRSVSPENLVKSRGSRGSPERSHVSSEPEYPISPQVPVHV--EKQDS 2164  
Db 2779 --ARVTFNY-----NPSPKSSADSTSARPSQIP--TPVNNNTKKRDS 2818

## RESULT 13

US-08-450-582-7

Sequence 7, Application US/08450582

Patent No. 6114124

## GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODIN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESS: Banner &amp; Witcoff, Ltd.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,582  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/452,655  
FILING DATE: 25-MAY-1995  
APPLICATION NUMBER: US 08/289,548  
FILING DATE: 12-AUG-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.49964  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
US-08-450-582-7

Query Match 2.9%; Score 365; DB 3; Length 2843;

Best Local Similarity 19.0%; Pred. No. 6.6e-14;

Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

QY 347 PETKQREQE-RFQVQAGLSATIASEH-----EISEIIDGLSEQENNEK 395  
Db 981 PSIESIEDDESKFCYGOYPADLAHKAHNSAMHMDNGELDPINYSLKYSDEQLNSGR 1040  
QY 396 QMRQLSVIPPMFDEARRVKFTNMGLMEDPMKYKDRQFMVWTDHE--KEIFKDKFI 453  
Db 1041 Q-----SPQNERWARPKHIIIDEIKQSEQRQSRNQSTPYVYTESTDDKHL 1087  
QY 454 QHPKFNGLIASYLERKSVPCVLYLYLTCKKNYKALVRNRYGKRRGRNQOIARPSQEEK 513  
Db 1088 KFQPHFG-----QQECVSPY---RSRGANGSETNRVGSNHGINQVQSLOCQ- 1132  
QY 514 VEKEEDKAKTEKK--EEKKDEEKEDEKSKENTKEKDKIDGTAEETEERQATPRG 571  
Db 1133 -DDYEDDKPTNYSERYSEEQHEEERTNYSIKYN-EERHVDQPIDYSLKYATDIPSS 1190  
QY 572 R-----KTANSQGRKGRITRSMTEAAAS-----AAAAATEEPPLPPP 614  
Db 1191 QKSFSSKSSSGSKSTEHWSSSESTSTPSSNAKQNLHPSSAOSRSGQPOKAATCK 1250  
QY 615 PEPISTPEVET-----SRTEEMEYAKKGLVEHGRN-----WAAIAKM 653  
Db 1251 VSSINQETIQYCVEDPICFSRCSLSLSAEDGICGQNTQEDASANTLQIAEIKK 1310  
QY 654 VTKSE-----AQCNFYNTKRRNLNLQOHK-----QKTSR 688  
Db 1311 IGRSADPDVSEVPAVSQHPRTKSSRLQGLSSSASARHRAVEFSSGAKSPKSGAATPK 1370  
QY 689 KPREE-----RVSOCESVASTVAQDEDEIEASNEEENPEDESEAVKPE--DSPEN 740  
Db 1371 SPPEHYVQETPLMFPSRCTSV--SSLDSFESRIASSVQSEPCSGMVGSIISDLPLDSPGQ 1429

QY 741 A--TSRGNTAPEVELEPTTETAPSTSPSLAVPSTKPAEDSESVETQVNDISISAETAQMDV 798  
Db 1430 TMPFSRSKTPPP---PPQATKREVPKNKAPTAKKESGPKQAANAQV--RVQVLPA 1485  
QY 799 DQOEHSAEE---GSVC-----DPPATKADSDVDVEVR-----VPEN-HASKVECDN 840  
Db 1486 DTLHFATSTPDGFCSSLSALSDEPFIQK-----DVELRIMPPVOENDNGNETESEQ 1541  
QY 841 TKE--RDLDRASEKVEPRDELVVAQINQARPERPQSDNDSSATCSADEVD----- 890  
Db 1542 PKESNEQEKAEKTIIDSEKDL-----DDSD-----DDIEIIEECIIS 1581  
QY 891 ---GEPERQRMFPMDSKPSLLNPTGSLVSSPLKNPLD---LP-----QLQRAAVIP 938  
Db 1582 AMPTKSSRKAKKPAQATASKLPPPVAR-----KPSQLPVYKLLPSQNRLQPOKHVSFT 1634  
QY 939 ---PMVSC---TPCNIPIGTPVSGYAL-----YQRHKAHES 970  
Db 1635 GDDMPRYCVEGTPINFSTATSLTIESPPNELAAGEVGRGAQSGEFEKRTDPTIEG 1694  
QY 971 AL-----LEEQROBOIDLECRSSSTPGCTSKSPNREWEVL----- 1007  
Db 1695 RSTDEAOGGKTSSVTIPELDDNKAEEDILAECLNSAMPKSKHKPRVKKIMDOVOQAS 1754  
QY 1008 --QPAPHQILTNLPEGVRLPTTRTPPPPLI-----PS 1039  
Db 1755 ASSAPNK---NQLDGKKKKTSPVKIPQNTYRTRVRKNADSKNNLNAERVSDNKDS 1811  
QY 1040 SKTVASEKFSF-----IMGSGIS-----QGTPTGYLTHNQOASTYQETPK 1080  
Db 1812 KQNLKNNSKDFNDKLPNEDRVGRSFAFDSPHHYTIEGTP--YCFSRND----- 1860  
QY 1081 PSVGSISLGLPROQESAKSATLPYIKOEFSRSONSOPELLVRAOHEGVVRGTAGAIQ 1140  
Db 1861 -SLSSLDFFDDVDLSREKAEKAKENKESAKVTSHTE--LFSNQOSA---NKTQAI 1914  
QY 1141 EGSITRGTPTSKISVESIPSLRGSITQGTALPOTGIPTEALVK-----GS 1186  
Db 1915 KQPINRGQPKPIQKQS-----TFQSSKDIIDRGAATDEKLFALIENTPVCFSHNS 1968  
QY 1187 ISRM-----PIEDSSPEKGRBEAASKGHVIEGKSGHI----- 1219  
Db 1969 LSSLSIDDOENNNKENEPKETETPPDSQGEPSKPKQASGYAPKSPHFVEDTPVCFSRNSLS 2028  
QY 1220 -LSVDN-----IKNAREGTRSPRTAHEISLKRKSVESVEGNIKQGMRESVPVSAPLE 1270  
Db 2029 SLSIDSEDDLQECISSAMPKPKPSRLKGDNEKSPNNMGILGEDLTLDKDIQRPDS 2088  
QY 1271 GLICRALPRGSPHSDLKERTVLGSGIMOGTPRATFESFEDGLKYPQIKRESPPIRAFEG 1330  
Db 2089 E--HGLSPDSENFDMKAIQEGANSIVSSLHQAAAAAC-----LSRQASSDSDSILSKS 2140  
QY 1331 AIYKGYDYGITTKEMGRSIEHEIPRODILTQESRKTPVVOVSTRPIEGSISOGTPIKF 1390  
Db 2141 GISLGSFP-----HLTPDQEEKPTTSNKGPRILK-----PGEKSTLETAKKI 2181  
QY 1391 DNNS-GQSAIKHNKVSITGP-----SKLSRGM--PPLBIIVPENIKVBERGKYEDVKAGETV 1444  
Db 2182 ESESKGLGKGVKVSILITGKVRNSNISQMKOPLQ---ANMPSISRG----- 2227  
QY 1445 RSRHTSVVSGPSVLRSTLHEAPKAQLSPGIYDTSARRTPVSYQNTMSRGSPMMRTSD 1504  
Db 2228 -----RUMIH-----IPGV-----RNSSSSTSPVSKKGPP 2252  
QY 1505 VTIPPNKSTNHERKSTLTPT--QRESIPAK--SPVPCVDVPSVSHSPFDPHHRGSGTAGEVWS 1562  
Db 2253 LKTPASKSPSEGQATTSRGAKPSVKSELSVARQTSQIGGSKAPSRSGR----- 2305  
QY 1563 HLTPLDPAAMPFFHRALDPAANAALFQRLSPTPGYPSQYOLYAMENTRQTLINDYITSOQ 1622  
Db 2306 -----DSTPSRAQOPL-----SRP 2320  
QY 1623 MQVNLRPDVA---RGLSPREQPLGLPYPATRGIIDLTMPPPTILVPHPGGTSTPMDRIT 1679

```
Db 2321 IOSPCRNISPGRNGISPPNK-----LSQLPT-SSPSTASTKSSGSKMS 2365
Qy 1680 YIPGTOTFPFPPPNYSMSGCHTHIAAASAEERERE-----KERERE 1727
Db 2366 Y-----TSPGRMSQOQLT--KOTGLSKNASSIPRSESASKGNOMNGANKKVELS 2417
Qy 1728 RIAAASDLYLRPQSGRPGRGSHGYVRSPSVRTOTMLQORPSVFOGTNGTSVITPL 1767
Db 2418 RMSSTKS-----GSESD-----RSERPVLRGSTFIKEAPS-----2449
Qy 1788 DPTAQLRIMPLPAGPSISQGLPASRYNTRADALAALVDAASAPOMDVSKTESKHEAA 1847
Db 2450 -PTLR-RKLEESAGFESLS---PSSRPASPTRSQAOPTVLSPLDMSLS-THSSVQAGG 2503
Qy 1848 --RLEENLRSAAYSCQOQLEQKTLVEKRSVOCULTSSAPFSGKQPPHSSVVYSAGK 1905
Db 2504 WRKLPNML---SPTIEYNDGRPAKRHDIAI-----SHSESPSRPLPTRNSGTWKREHSK 2553
Qy 1906 DKGPPPKSRYEELRTRGKTTITAN-----FIDVITRQJASDKDARERGSSQSDS 1957
Db 2554 HSSSLP--RVSTWRTGSSSILSASSESEKAKSEKHKVNSISGTKQSKENOVSAKGT 2611
Qy 1958 SSSLSHRYE-TPSDAIEVISPASPAPQOEKLTQYQEVVYKANOENDPTRQYEGPLHH 2016
Db 2612 WRKIKENEFSTNSTQTVSSGATNGAESKTLIYOMAPAVSK---TEDVMVRIEDCPINN 2668
Qy 2017 YRPOEESPSQOQLPP-----SSQAGMCOVPRTHRLITLADHIC 2056
Db 2669 PR---SGRSPTGNTPPVIDSVSEKANPNIKDSKDNQAKQNGVSPM--RTVGLNRLN 2723
Qy 2057 QIITQDFARNQVSSQTP--QOAPTSTFQNSPVALYS--TPVRTKTSNRYSPESAOQSVHHQR 2114
Db 2724 SFIQVDAPDQKTEIKQCONNPVPVSETNESSIIVERTFFSSSSSKHSSPGTVA-----2778
Qy 2115 PGSRVENLVDSKGRSGRPSKPSRSHVSEPEYPISPPOVPPVHH--EKQDS 2164
Db 2779 --ARVTFPNY-----NPSRKSSADSTSARPSQIP-TPVNNNTKKRDS 2818
```

## RESULT 14

US-08-821-355A-7  
Sequence 7, Application US/08821355A  
Patent No. 5851775

## GENERAL INFORMATION:

APPLICANT: Barker, Nick  
APPLICANT: Clevers, Hans  
APPLICANT: Korinek, Vladimir  
APPLICANT: Morin, Patrice  
APPLICANT: Kinzler, Kenneth  
APPLICANT: Vogelstein, Bert  
APPLICANT: Sparks, Andrew  
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC  
TITLE OF INVENTION: Interact to Prevent Cancer  
NUMBER OF SEQUENCES: 11

## CORRESPONDENCE ADDRESS:

ADDRESS: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, N.W.  
CITY: Washington

## STATE: DC

## COUNTRY: USA

## ZIP: 20001

## COMPUTER READABLE FORM:

## MEDIUM TYPE: Diskette

## COMPUTER: IBM Compatible

## OPERATING SYSTEM: DOS

## SOFTWARE: FASTSEQ for Windows Version 2.0

## CURRENT APPLICATION DATA:

## APPLICATION NUMBER: US/08/821,355A

## FILING DATE: 20-MAR-1997

## CLASSIFICATION: 514

## PRIOR APPLICATION DATA:

## APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A  
REGISTRATION NUMBER: 32,145  
REFERENCE/DOCKET NUMBER: 1107.05064  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 97430 BMB UT  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2973 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5851775e  
US-08-821-355A-7

Query Match 2.9%; Score 365; DB 2; Length 2973;  
Best Local Similarity 19.0%; Pred. No. 7.1e-14;  
Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

```
Qy 347 PEIKQREOQE-RFQRYGQRGAGLSATIRSEH-----EISEIIDGLSEQENNEK 395
Db 981 PSIESYSEDDSKFCSYGQYPADLAHKIHSANHMDNDGELDTPIYSLKYSDEQLNSGR 1040
Qy 396 QMRQLSVIPPMFPAEORRYKFINNGLMEDPMKYKDRQFMNVWTDHE--KEIFKDKFI 453
Db 1041 Q-----SPQNERWARPKHIIIEIKQSQRSRQNSTTYPVYTESTDDKHL 1087
Qy 454 QHPKNFGLIASYLERKSVDPCLVYLYLTKNENYKALVRRNYGKRRGRNOQIARPSOEK 513
Db 1088 KQPHFG-----QQECVSPY----RSRGANGSETNRVGSNHNQINQNVSSLCQE- 1132
Qy 514 VEEKEEDKAEKTEKK--EEKKEDEEKEDEKEDSKENTKDKIDGTAEETEEREQATPRG 571
Db 1133 -DYEDDKPTNYSERYSEEQHEEERPTNYSIKYN-EKRRHVDQPIDYSLKYATDIPSS 1190
Qy 572 R-----KTANSQGRKGRITRSMNTNEAAS-----AAAAAATEEPPPLPPP 614
Db 1191 QKOSFSSKSSGSSKTEHMSSESTSTPSSNAKQNLHPSSAQSRSQGPQKAATCK 1250
Qy 615 PEPITSEPVET-----SRWTEEMEYAKKGLVEHGRN-----WAAIAKM 653
Db 1251 VSSINQETIQTICVEDTPICFSRCSLSLSAEDGICNQTQEADSAANTLQIAEIKK 1310
Qy 654 VGTKSE-----AOCKNFYFNKRRNLDNLQOHK-----QKTSR 688
Db 1311 IGTRSAEDPVSEVPASVQHPRTKSSRLQSSLSSESARHKAVERFSGAKSPSKGAQTPK 1370
Qy 689 KPREE-----RQVSOCSVASTVSAQEDEDIASNEENPEDESEVAVKPE--DSPEN 740
Db 1371 SPPEHYVQETPLMFSRCTSV--SSLDSEFSIASVSSSEPCSGMVSGIISPSDLPSPGQ 1429
Qy 741 A--TSRGNTEPAVELEPTTETAPSTPSLAVPSTKPAEDSESVETQVNDSDISAETAQMDV 798
Db 1430 TMPPSRSKTPPP---PPQATQREVKNKAPTAKEKESGPKQAANVAQV-RVQVLPDA 1485
Qy 799 DQOEHSAAE-----GSVC-----DPPPATKADSVDEVR-----YPEN-HASKVGEDN 840
Db 1486 DTLLHFATESTPDGFCSSSLSALSALDEPFQK---DVELRIMPPVQENDNGNETESEQ 1541
Qy 841 TKE--RDLDRASEKVEPRDEDLVVAQOINAQRPQSDNDSSCATSADSDVD-----890
Db 1542 PKSENENQEKAEKTIIDSEKDLL-----DDSD-----DDDIIELEECIIS 1591
Qy 891 ---GEPERQRMFPMDSKPSLLNLTGTSILVSSPLKPNPLD----LP-----QLQRAAVIP 938
Db 1582 AMPTKSSRKAKKPAQATASKLPPPVAR-----KPSOLPVYKLLPSQNRLOPKQHVFTFP 1634
Qy 939 ----PMVSC---TPCNIPITGTPVSYAL-----YORHIKAMES 970
```

Db 1635 GDMPRVYCVETGPIINFSTATSLDLETSPNLAAGEVRGGAQGEFEKRDITPTIG 1694  
Qy 971 AL-----LEORQOEIDLECRSSSPCOTSKSPNREWEVL----- 1007  
Db 1695 RSTDEAOGGKTSSVTIPELDNKAEEGDILAECINSAMPKSKHPFKVKIMDOVOQAS 1754  
Qy 1008 --QPAPHOLITNLPPEGVRLPTRRPPPLI-----PS 1039  
Db 1755 ASSAPNK--NQLDGKKKPTSPVKPIQONTYRTRVRKNADSKNLAERFVSDNKDS 1811  
Qy 1040 SKTIVASEKPSF-----IMGGSIS-----OGTPGYLTHSHNOASTQETPK 1080  
Db 1812 KQQLKNNSKDFNDKLPNNEDRVGRSFAFDSPHHYTPIEGTH--YCFSRND----- 1860  
Qy 1081 PSVGSISLGLPQOESAKSATPYIKOEFSPRSONSOPEGLLVRAOHEGVVVRGTAGAIQ 1140  
Db 1861 -LSLSLDFDDDDVLSREKAEKAKENKESEAKVTSHTS--LTSNQOSA---NKTQAI 1914  
Qy 1141 EGSTRTPTSKISVESIPSLRGSITOGTPALPOTGIPTEALVK-----GS 1186  
Db 1915 KOPINRGOPKPILOKS-----TFQSSKDIPDKGAATDEKLQNFALIENTPVCFSHNS 1968  
Qy 1187 ISRM-----PIEDSSPEKGRREAAKGHVIEKSGCHI----- 1219  
Db 1969 LSSLSIDIOENNKNENEFIKETEPDPSQGEFKPQASGYAPKSFHVEDTPVCFGRNSLS 2028  
Qy 1220 -LSYDN-----IKNARECTSPRTAHEISLKRVSYESVEGNKQGMRESVPSAPLE 1270  
Db 2029 SLSIDSEDDLLQECISSAMPKKPSRLKGDNEKHSRPNMGIIIGEDLTDLKDQRPDS 2088  
Qy 1271 GLICRALPRGPHDLKERTVLGSIIMQGTTPRATESFEDGLKYPKQIKRESPIRAFEG 1330  
Db 2089 E---HGLSPDENFDKAIQEGANSIVSSLHQAAAAAC-----LSRQASSDSLSILKS 2140  
Qy 1331 AITKGKPYDGTITIKEMGRSHEIPRODILTQESRKPEVVOSTRPIEGSISOGTPIKF 1390  
Db 2141 GISLGSFP-----HLTPDQEEKPFTSNKGPRIK-----PGEKSTLETKKI 2181  
Qy 1391 DNNS-GOSAIAKHNKSLITGP-----SKLSRGM-PPLEIVPENIKVBERGKYEDVKAGETV 1444  
Db 2182 ESESGIKGGKVKYLSITKVRNSNLSISOMKQPLQ---ANMPSIRG----- 2227  
Qy 1445 KSRHTSVVSGPSVLRSLTHLEAPKAQLSPGIYDDTSARRTPVSYQNTMSRSPMNRSTD 1504  
Db 2228 -----RTMIH-----IPGV-----RNSSSSTSPVSKGKGP 2252  
Qy 1505 VTIPPKNSTNHERKSLTPT-QRESIPAK-SPVPGVDPVVSHPDFDHHRCSTAGEVYWS 1562  
Db 2253 LKTPAKSPSEGOATTSPRGAKPSVKSELSPVARQTSQIGGSKAPSRSGR----- 2305  
Qy 1563 HLPQOLDPAMPFHRALDPAAAAYLFQRLSPTPGYPQYQOLYAMENTROTILNDYITSQ 1622  
Db 2306 -----DSTPSRPAQOPL-----SRP 2320  
Qy 1623 MQVNLRPDVA---RGLSPREOPLGLPYATRGIIIDLNMPTILVPHPGSTPMPDRIT 1679  
Db 2321 IQSPGRNISPGRNGISPPNK-----LSQLPRT-SSPSTASTKSSGSGKMS 2365  
Qy 1680 YIPCTQITFFPRPNYSASMSGPHTHLAAASAERERERE-----KERERE 1727  
Db 2366 Y-----TSPGRQMSQNLIT--KOTGLSKNASSIPRSEASKGLNQMNNGANKKVELS 2417  
Qy 1728 RIAAASDLXLRPGSEQPGRPGSHGVYRSVSPSVRTQETMLQQRPSVFGQNGTSVITPL 1787  
Db 2418 RMSSTKSS-----GSESD-----RSERPVLVRSTFIKEAPS----- 2449  
Qy 1788 DPTAQLRIMPLPAGGPIISQGLPASRYNTAADALAAALVDAASAPQMDVSKTESKHEAA 1847  
Db 2450 -PTLR-RKLEESASFELS---PSSRPASPTRSQAQTPVLSPLDMSLS-THSSVQAGG 2503  
Qy 1848 --RLEENLRSEAAVSEQOQLEQKTELEVERKSVOCLTSSAFPSGKQPQSHSVVYSAGK 1905  
Db 2504 WRKLPPNL---SPTIYNDGRPAKRHDIAI-----SHSESPRLPINRSGTWKREHSK 2553

Qy 1906 DKGPPKSRYSREELRTRGKTTITAN-----FIDVILITROIASDKDAREGSSSDS 1957  
Db 2554 HSSSLP--RVSTWRTGSSSILSASSESEKSEKSEKSEKSEKSEKSEKSEKSEKSEK 2611  
Qy 1958 SSSLSHRYE-TPSDAIEVISPASPAPQPKLOTYQEVVYKANAENDPTRQYEGPLHH 2016  
Db 2612 WRKIKENEFSTNSTQTVSSGATNGAESKTLIYQMAPAVSK---TEDVWVRIEDCPINN 2668  
Qy 2017 YRPOESPSPOOQLPP-----SSQAEQMGQVPRTHRLITLADHIC 2056  
Db 2669 PR---SGRSPTGNTPPVIDSVSEKANPNIKDSKONAKQNGVSPM--RTVGLENRLN 2723  
Qy 2057 QIITQDFARNQVSSQTP--QOPPTSTFQNSPVALYS-TPVTRKTSNRYSPESQAOQSVHHQR 2114  
Db 2724 SFIQVDAPDQKGTETKPCQNNPVVSETNESSIVERTPFSSSSSKHSSPSGTV 2778  
Qy 2115 PGRSVSPENLVKSRGSRGKSPERSHVSSEPFISPPQVPPVH--EKQDS 2164  
Db 2779 --ARVTPFN-----NPSRKSSADSTARPSPQIP-TPVNNNTKKRDS 2818

## RESULT 15

US-09-003-687A-7  
; Sequence 7, Application US/09003687A  
; Patent No. 5998600  
; GENERAL INFORMATION:  
; APPLICANT: Barker, Nick  
; APPLICANT: Clevers, Hans  
; APPLICANT: Korinek, Vladimir  
; APPLICANT: Morin, Patrice  
; APPLICANT: Kinzler, Kenneth  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Sparks, Andrew  
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC  
; INTERACT TO PREVENT CANCER  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/003,687A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/821,355  
; FILING DATE: 20-MAR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A  
; REGISTRATION NUMBER: 32,145  
; REFERENCE/DOCKET NUMBER: 1107.05064  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; TELEX: 97430 BMB UT  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2973 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5998600e  
US-09-003-687A-7





```
Qy 1728 RIAAASDLYLRLPSEQGRPGSHGYVRSPSPSVRTQETMLQQRPSVFQGTNGTSVITPL 1787
Db 2418 RMSSTKSS-----GSESD-----RSERPVLVRQSTFIKEAPS----- 2449

Qy 1788 DPTAQLRIMPLPAGGPSISQGLPASRYNTAADALAAALVDAASAPQMDVSKTKESKHEAA 1847
Db 2450 -PTLR-RKLEESASFESLS---PSSRPASPTRSOAQTPVLSPSLDMSLS-THSSVOAGG 2503

Qy 1848 --RLEENLRSAVSEOOQLEQTELEKESVOCCLYTSSAFPSGCKPQPHSHSVVYSEAGK 1905
Db 2504 WRKLPPNL---SPTIYNDGRPARHDIAR-----SHSESPRLPNRSGTWKREHSK 2553

Qy 1906 DKGPPPKSRYEEELRTKTTITAA-----FIDVIITQIASDKDAREGQSDDS 1957
Db 2554 HSSSLP--RVSTWRTGSSSILSASSESEKAKSEDEKHVNSISGCKQSKENQVSAKT 2611

Qy 1958 SSSLSHRYE--TPSDAIEVISPASPPQEKLOTYOPEVVKANOANDPTROYEGPLHH 2016
Db 2612 WRKIKENEFSTNSTQTVSSGATNGAESKTLIQMAPAVSK---TEDVWVRIEDCPINN 2668

Qy 2017 YRPOQESPSPOOLPP-----SSQAEQMGQVPTHLITLADHIC 2056
Db 2669 PR-----SGRSPGTNPVVIDSVSEKANPNIKDKNQAKQNGVNGVPM--RTVGLENRLN 2723

Qy 2057 QIITQDFARNQVSSQTP--QOPPTSTFQNSPALSYS--TPVRTKTSNRYSPESQAQSVHQR 2114
Db 2724 SFIQVDAPDQKTEIKPQNNPVPVSETNESSIVERTPFSSSSSKHSSPSGTVA----- 2778

Qy 2115 PGRSVSPENLVKSRGSRGKPSRSHVSSPEPIPSQVPPVHH--EKQDS 2164
Db 2779 --ARVTFPNY-----NPSRKSSADTSARPSQIP-TPVNNNTKKRDS 2818

RESULT 17
US-08-061-376-5
; Sequence 5, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Djabali, Malek
; APPLICANT: Selleri, Lucia
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
; TITLE OF INVENTION: TRANSLATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: 941 9387
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)546-4737
; TELEFAX: (619)546-9392
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3969 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
```

```
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-061-376-5

Query Match 2.6%; Score 329; DB 4; Length 3969;
Best Local Similarity 17.5%; Pred. No. 1.8e-11;
Matches 593; Conservative 378; Mismatches 1089; Indels 1332; Gaps 153;

Qy 1 MSSSGYPPNQAGAFTEOSRYPPHVSQVTFPTNRRHQQEFV-----PDYRSHL 48
Db 278 LSPSSLSLROGSEFKZEGRYKLTZTERKASINRKDKDFGLLIILNWKSPRKSGTKREHL 337
Qy 49 E-----VSOASQLLQOQOOLRRRPSLLSEFHPGSDRPOBRSYTPPFH 93
Db 338 HLQKKIRQLSDKALEGLSQLGLFLQKQMO-----PLLSNSYRGQKKGAQKKEKA-- 390
Qy 94 PGSPVDHDSLESKRPRLEOVSDSHFORVSAAVLPLVHPLPEGL-----RASADAKDP- 147
Db 391 -----AQLQGRKVK-----TQVKNIRQFIMPVVSATSSRIKTPRRFIEDYDPP 436
Qy 148 -AFGKKHEARSSPISGQPCGDDQNAS-----PSKLSKEELIQMDRVDRERET---AKVEQQ 198
Db 437 IKIARLESTPNRSFAPSACSGSEKSSAASQHSQMSDSSRSSPSVDITSDSOASEIQ 496
Qy 199 IL-----KLNKKQOOLEEEA 213
Db 497 VLPERSDTPVHPPLPISQSPENESNDRRRYSVRSERFSGSRTTKLSTLQSAPOQT 556
Qy 214 AK-----PPEPEKPV-----PP-----PVEQKHSIV----- 236
Db 557 SSSPPPLLTTPPLQPASSISDHTPWLMPPTIPLASFPPLASTAPMQGKRKSILREPTF 616
Qy 237 -----QIYDENRKAEEA---HKIFEGLPKVELPLYNQPSDT----- 272
Db 617 RWTSLKHSRSEPQYFSSAKYAKEGLIRKPIFDNFRPP---PL--TPEDVGFASGFSAGT 671
Qy 273 -----KVYHENIKTNQVMRKLI---LFFKRNHARKQREKICORYDOLMEAWKKV 322
Db 672 AASARLPSPLHSGTRFDMHKRSPLLRAPRTTPSEASR-----IFESVILPS 718
Qy 323 DRINNPRKAKESKTREYEEKQPEIRKQEQOERFORVQORGAGLSATIAREHEISE 382
Db 719 NRTSAGTSSSGSVNRKRK--RKVESPIRSPSPSHSMRT--RSGRLSSSELSLTPTPSS 774
Qy 383 IIDGLS-----EQENNEKQMRQLSVLPMPMFAEQRRV 415
Db 775 VSSLSISVSPPLATSALNPTFTFPFSLTSGESAENQRPRKOTSA-PAEPFSSSSPTP 833
Qy 416 KF-----INMGLMED--PMKVYKDRQFMNVMWTDHEKEIFKDKF----- 452
Db 834 LFPWFTPGSQTERGRNKKDCAPELSKDR-----DADKSVEKDKSRDREREKENRE 886
Qy 453 -----IQHPKNFGLIASYLERKSVDPDCVLYLYLTKNE----- 485
Db 887 SRKEKRKKGSEIQSSSALYPVGRVSKYKGVGEDVATSSAKKATGRKKSSHSDSGTDITS 946
Qy 486 -----NYKALVRNYKGRGRNQOIARPSQEEKVEEK----- 517
Db 947 VTIGDTTAVTKILIKKGRGNLEKTNLDLG--PTAPSLKEKTKLCLSTPSSSTVHKHSSI 1005
Qy 518 -----EEDKAEKTEKKE-----EEKKDEEEKDEKSKENTKEKIDGTAEETEEREQAT 568
Db 1006 GSMLAQADKLPMWTDKRVASLLKKAQCLCKIEKSKLQOT--DQPKAQQESDSSETS 1064
Qy 569 PR-----GRKTANSQGRKG-----RITRSWTNEAAASAAAAATEE 606
Db 1065 PRIKHVCRAAVALGRRAVFPDMDPTLSALPWPEREKILSMGNDKSSTAGS-----E 1119
Qy 607 PPPPLPPPPISTEPVETSRWTEEMEVAKKGLVEHGRNWAALAKMVGTKSEAQ----- 662
Db 1120 DAEPLAPPKPI--KPVTRNKAPQEP-----VKGRSRRCGCGCGVPCGVCCT 1170
```





Db 3089 -----VLQTLPLNGVTKIQLTSSVSTPSPVMTNTSV---LGPWGGGLTLTLGLNP---- 3136  
QY 2235 IFNLPAVTTSGSVSRG-----HSPADPASNLGLEDIIIRKALMGSDFDKVEDHGV 2284  
Db 3137 --SUPTOSLPSASKGLLPSHSHQHLSF--PA-----ATQSSPPPNIS----- 3177  
QY 2285 VMSOPMGVPGT---ANTSVTSGETREED---PSPHSGGVCKPKLISKSRSRKSXP 2338  
Db 3178 --NPPSGLLIGVQPPDQLLVSESSQRTDSTVATPSSG--LKKRPISLRQTKKKKL 3233  
QY 2339 IPGOGYLGTERPSV--SSVHSEGDYHROTPGMAWEDRPSSTGSTQPPYNPPLTMRMLSS 2396  
Db 3234 AP-----SSTPSNIAPSDVSNMTLNFTP-----SQLPNHP-SLLDLGSL 3273  
QY 2397 PPTIACAPSAVNAAPHOQNRINEREPAPLL 2428  
Db 3274 NTSSHRTVPNIKRS----KSSIMYFEPAPLL 3301

RESULT 18  
US-08-769-309A-5  
; Sequence 5, Application US/08769309A  
; Patent No. 5741890  
; GENERAL INFORMATION:  
; APPLICANT: Scott, John D.,  
; APPLICANT: Nauert, Brian J.,  
; APPLICANT: Klauck, Theresa M.,  
; TITLE OF INVENTION: Protein Binding Domains of Gravin  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower/233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/769,309A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5741890and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 27866/33451  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1780 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-769-309A-5

Query Match 2.6%; Score 323.5; DB 1; Length 1780;  
Best Local Similarity 18.6%; Pred. No. 1.3e-11;  
Matches 361; Conservative 273; Mismatches 741; Indels 565; Gaps 82;

QY 496 KKRGRNQIARPSQEKV-----BEKEDKAEKTEKKEEKKDEKDEKDSKENT 548  
Db 83 GQKALNGOGALNSQEEIEVTVGQDSSEDSERDSKEMATKSAVVHDIITDQGEEN 142  
QY 549 KEKDKIDGTAEETEEREOQAPRGRKTKANSOGRRK-----GRITR 587  
Db 143 RNIEQIPSSSENLEELTQPT---ESQANDIGFKKVFVFGFKTVVKDKTEKPTVOLLT 199

QY 588 SMTNEAANAASAA-----AAAATEPPPLPPPEPISTE-----PVETS 627  
Db 200 VKDEGGAAGAGDHQDPSLGAEEAKSEKPEKOSTEKPEETLKREQSHABISPAESGQ 259  
QY 628 WTEF-EMEAVAKGLVHGRNAAIAKVMGTSKSEAQCNFYFNKRRHNLNLLQHQKQT 686  
Db 260 AVECKEKEGEKQEKESKSAESPTSPVTSSTGTFKFF-----TQWAGWRKTS 311  
QY 687 SRKPREDRVSQESVASTVSAQDEDEDIEASNEENPEDSEVEAVKPSDESPENATSRGN 746  
Db 312 FRKPE-----DEVEASEKKQEPEKVDTERDGAEVAASEKLTAEOAHQPOEPAESA 364  
QY 747 TEP-----AVELEPTTETAPSTPSLAVPSTPAE--DESVE----- 781  
Db 365 HEPRLSAEYKVELPSEQVSGSQGPSEKPAPLATEVFDEKIEVHQQEVEVVAEYHVHSTVE 424  
QY 782 -----TQVND---SISAETAEMDVDOEH-----SAEGSVCDPPPPATKAD--SVD 823  
Db 425 ERTEEQKTEVEETAGSVPAEELVGMDAEPQEAEPKELVKLKEKTCVSGEDPTQCADLSPD 484  
QY 824 VEV--RVPENHASKVEGDNTKER-----DLDRASEKVEP-----RDEDLV 861  
Db 485 EKLSKPEGVVSEVEMLSQERMKVQGSPLKLFSTGLKLLSKKQKGRGGDEESG 544  
QY 862 VAOQINAORPEQSDNDSSATCSADE-----VDEPERQ 896  
Db 545 EHTQVPADSPSQEEQKGESSASSPEPEITCLEKGLAEVQOGEAEBGATSDEKRE 604  
QY 897 RMFMDSKPLLNPFTGSLVSSPLKNPLDLQLQHRAAVIPPMVSVCTPCNIPITGVSG 956  
Db 605 GVTPWASFKKWTP----- 618  
QY 957 YALYORHIKAMHESALLEBQRQEQIDLECRSTSPCTGTSKSPNREWEVLQAPHQLIT 1016  
Db 619 -----KKVRPPSES-----DKEDLDKVKSATLSSTESTASEMOEEMKSGVEEPAPEE-- 667  
QY 1017 NLPEGVRLPTRTRPPPLIPSPKTTVAESKPSFIMGSSISQGTGTYLTSHNQASYQ 1076  
Db 668 -----PRKVDTSVWEALICVSGSKK-----RARRSSSDE 699  
QY 1077 ETPKPSVGSISLGLPROQESAKSATLPYIKOEFPSPQNSQPEGLVLVRAHQHEGVVGT 1136  
Db 700 EGGPKAMG-----DHQRADE-AGDKKETGTDGILAGSOEHDPGGQSS 741  
QY 1137 GATQEGSITRPTSKISVESISLRSITQGTALPQTPQTIPTALYKGSISRPIEDSS 1196  
Db 742 SPEQAGS-----PTEGEGVSTWESFKRLVTPRKSKSLEKSEDSSTAGS-----GVEHST 792  
QY 1197 P--EKGREEAASKGHVIEGKSHLSYDNIKNAKREGTRSPRTAHEISLKRYSVEGNI 1254  
Db 793 PDTEPGKEE-----SWVSIKKFIPORRKRP-----DGQEQAPVEDAG 831  
QY 1255 KQMSMRSPVSA--PL-----EGLICRALPRGSPHSDLKERTVLSSIMQGTPTAT 1304  
Db 832 PTGANEDSDVPVAVVPLSEYDAVEREKMEAQAQQAQEAQEAQEAQEAQEAQEAQEAQEA 891  
QY 1305 TESFEDGLKYPKOIKRESP-----PIRAF--EGAITKGPYDGTITIKEMGRSIEH 1353  
Db 892 AAADVADGTRAATIEERSPSWISASVTEPLEQVEAEALITEEVLEVEVIAEETPTVTE 951  
QY 1354 -IP-----RODILTOESRKTPEVVQVSTRPIEGSISQGTPIKFDNNSGQSAIKHN----- 1402  
Db 952 PLPENREARGDVTVSEAEALTPEAVTAA-----ETAGPLGSEEGTEASAAEETEMVS 1003  
QY 1403 -VKSITGPKSLSRGMPLEI---VPENIKVVERGKVEDKA-GETVRSRHTSVSSGSP- 1456  
Db 1004 AVSOLTDSPDPTTEATPVQEVGGVD-IEEQERRTOEVLQVAEAKVKEESQLPGTGPE 1062  
QY 1457 SVLRSTLH---EAPKAOL-SPGIYDDT-----SARRTPVSYQNTMRSGSP-MMRT 1502  
Db 1063 DVLQPVORAEAREPEQEAESGLKKEKTDVVLKVDQAQEAKEPTTQGVQVGTTPESFEKA 1122

Qy 1503 SDVT--IPPNKSTNHERKSTLTPTQRESIPAKSPVPGVDPVSHSPFPHRGSTAGEVY 1560  
Db 1123 PQVTIESSELVTTQCAETLAGVQSBWMEQAIP--PDSVETPTDSETDGTTPVADF 1179  
Qy 1561 WSHLPTQLDPAMPFHRA-----LDPA-----AAAYLFQROLSTPTGYPQYQIYAMENTROT 1612  
Db 1180 DAPGTTQKDEIVEHEENHVLVPRGTAEAVPAQKERPPA---PSSF-VFQETKREQS 1235  
Qy 1613 ILNDYITTSQQMVNLPRDPAVAGLSR-----EPLGLP-YPATRGIDLTNNPP 1660  
Db 1236 KMEDTLEHTDEKVS--ETVSILSKTCTEQADQYADEKTKDVFFEGLEGSID-TGI-- 1290  
Qy 1661 TILVPHPGGTTPTPMDRTYI-----PGQIITFPR-----PYNASMSPHGPHTHLAAAASA 1712  
Db 1291 -----TVSREKVTALVAGEETAEACKDDALELQSHAKSPSPVEREMVQV 1339  
Qy 1713 ERER-ERE-----REKERERERIAAASDLYLRPGSEQ-----PGRGSHGYVRSP 1757  
Db 1340 BREKTEAEPHNEEKLHETAVTVSEEV-----SKQLLOTVNVPIIDGAKEVSSLEGSP 1394  
Qy 1758 SPVVRTQETM-----LQORPSVFQGTNGTSVITPLDPTAQLRIMPL-----PAGGPSISQ 1807  
Db 1395 PPLCQOEAVCTKIQVQSSSEASFILTAAREEKKVLGETA--NILETGETLEPAGAHVLVE 1452  
Qy 1808 GLPASRYNTAA-----DALAALVDAAASAPQMDVS-----KTKESKHEAAR 1848  
Db 1453 EKSEKNEEDFAHPGEDAVPTGPDCAKSTPIVYSATTKGLSSDLEGEKTTSLKWSDE 1512  
Qy 1849 LEENLRSAAVS-----EQOQLEQKYLEKRS-----VOCUYTSSAPPSG 1890  
Db 1513 VDEQVACQEVKSVVAIEDLEPENGILELETKSKLVQNIQTAVDQFVTRTEATEMLTS 1572  
Qy 1891 KPQHSVVYSEACKGPPPKRYEB--ELTRGKTTITTAANFIDVITRQIASDKDARE 1949  
Db 1573 ELQTAHVIKADS--QDAQETEKGEPEQASQAQETPITSA----- 1612  
Qy 1950 RGSQSSSSSLSHRYETPDAIEVISPASSPAPPQKLOTQYQVYVYKANAENDPTRQ 2009  
Db 1613 --KEESTAVGAH-----SDISKMSSEAS-----EKTMTVEVE-----GSTVND--QQ 1653  
Qy 2010 YEGPLHYRPOQSPSPQOQLPPSSQAEQMG--QVPRTHRLITLADHICQIITO----- 2061  
Db 1654 LE-----EVLWPSSEEGGAGTKSVPEDDGHALLAERIEKSLVPEPKEDEK 1698  
Qy 2062 ----DFARNQVSS-----QTPQOPPTSTFONSPS-----ALVSTPVRTKTS 2098  
Db 1699 GDDVDDPENQNSALADTDASGLTKESPD-----NGPKQKEKEDAQEVLEQEGKVHSESD 1754  
Qy 2099 NRYSPESQAQSVHHRPGRS 2118  
Db 1755 KAITPQAQEELOKQERESAK 1774

## RESULT 19

US-08-994-570-5

Sequence 5, Application US/08994570

Patent No. 6090929

GENERAL INFORMATION:

APPLICANT: Scott, John D.,

APPLICANT: Nauert, Brian J.,

APPLICANT: Klauck, Theresa M.

TITLE OF INVENTION: Protein Binding Domains of Gravin

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESS: Marshall, O'Toole, Gerstein, Murray &amp; Borun

STREET: 6300 Sears Tower/233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/994,570

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: No. 6090929and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 27866/33451

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1780 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-994-570-5

Query Match

Best Local Similarity 2.6%; Score 323.5; DB 3; Length 1780;

Matches 361; Conservative 273; Mismatches 741; Indels 565; Gaps 82;

Qy 496 KGRGRNQIARPOEKEV-----EKEEDKARKTEKKEEKEDEKEDKESKENT 548  
Db 83 GQKALNGQGALNSQEEVIVTEVQGRSDSEVSDERSDKEMATKSAVVHDTDDQGEN 142  
Qy 549 KEKDKIDGTAEETREQATPRGRKTANSQGRK-----GRITR 587  
Db 143 RNIOIIPSSSENLELTQPT---ESQANDIGFKVKFKVGVFKVTKTEKPDVTQLLT 199  
Qy 588 SMTNEAAAAA-----AAAATEPPPPPLPPPPPEPISTE-----PVETSR 627  
Db 200 VKDEGEAGAGDHDPSLAGEAASKESEPKSTKEPTLKRQSHAEISPPAESGQ 259  
Qy 628 WTER-EMEVAKKGLVEHGRNWAATAKMWGTSEACKNFYFNYKRRHNLNLLQOHKOT 686  
Db 260 AVECKEKEGEKQKESKSAESPTSPVTSSTGTFKFF-----TQWAGWRKKTS 311  
Qy 687 SRKPREERDVQCSVASTVSAQEDIEDIASNEENPDSEVAVKPSEDESPENATSRN 746  
Db 312 FRKPE-----DEVEASEKKKQEPKVDTEEDGKAEVASEKLTASEQAHQPAESA 364  
Qy 747 TEP-----AVELEPTTETAPTSPSLAVPSTKPAE--DESVE----- 781  
Db 365 HEPRLSAEYKVELPSEQVSSQGPSEKPAPLATEVFDEKIEVHQEVEVAEVHVSVE 424  
Qy 782 -----TQVND---SISAEIAQMDVDQEH-----SAGESSVCDPPPATRAD-SVD 823  
Db 425 ERTEQKTEVEETAGSVPAEELVGNDAEPQAEPAKELVCLKETCVSGEDPTQCADLSPD 484  
Qy 824 VEV--RVYPENHASKVEGDNTKER-----DLDRASEKVEP-----RDELIV 861  
Db 485 EKVLSPPEGVVSEVEMLSQRMKVQGSPLKLTSTGLKLKSKKOKKGRGGGDEESG 544  
Qy 862 VAQINAQRPFPQSDNDSSATCSADE-----VDEPERQ 896  
Db 545 EHTQVPADSPDSQEEQKGESSASSPEEPTITCLEKGLAEVQDGEAEAGATSDGEKKRE 604  
Qy 897 RMFPMDSKPSLLNPTGSLVSSPLKPNPLDLPQLQHRAAVIPPMVWVCTPCNIPICTPVSG 956  
Db 605 GVTWASFKKMVT----- 618  
Qy 957 YALYORHIKAMHESALLEEQROEQIDLECRSSTSPGTSKSPNREWEVLQAPHOLIT 1016  
Db 619 -----KKVRPSES-----DKDELDKVKASATLSSTESTASQMOEEMKGSVEEPKEE--- 667  
Qy 1017 NLPEGVRLPTTRTPRPPLIPSSKTTVASEKPSFIMGSGISQGTGTYLTSHNQASYTQ 1076

```
Db 668 -----PKRKVDTSVSWEALICVSSKK-----RARRRSSDE 699
Qy 1077 ETPKPSYSGISLGLPROESAKSATLPYIKOEFSRSONSQPGLLVRAQHEGVVGT 1136
Db 700 EGGPKAMG-----DHQKADE-AGDKETGTDGILAGSQEHDPGGSS 741
Qy 1137 GAIQEGSITRTPTSKISVESIPSLRSGITQTPALPQTGIPTEALVKGSISRMPIEDSS 1196
Db 742 SPEQAGS-----PTEGEGVSWESFKRLVTPRKSKSLEKSEDSIAGS-----GVEHST 792
Qy 1197 P--EKGREAAASKHVIEGSKHILSYDNINKNAREGTRSPRTAHEISLKRYSVESGNI 1254
Db 793 PDTEPGKE-----SWVSIKKFIPIGRKKRP-----DGKQEQAPVEDAG 831
Qy 1255 KQGMRESVPVA--PL-----EGLICRALPRGSPHSDLKERTVLGSGIMOGTPRAT 1304
Db 832 PTGANEDSDVPVAVPSEYDAVEREKWEAQOAGAEQPKAATEVSKELSESQVHMM 891
Qy 1305 TESFEDGLKYPKQIKRESP-----PIRAF--EGAITKGPYDGTITTIKEMGRSIEH 1353
Db 892 AAADVADGTRAATIIERSPSWISASVTEPLEQVEAEALLTEEVLEEVIAEEPPVTVE 951
Qy 1354 -IP-----RQDILQESKKTPEVQSTRPIIEGSIQGTPIKFNNSQSAIKHN----- 1402
Db 952 PLPENREARGDTVVSEAELETPAVTAA-----ETAGPLGSEEGTEASAAEETTEMVS 1003
Qy 1403 -VKSILITGPSKLSRGMPLLET---VPENIKVVERGKYEDVKA-GETVRSRHTSVVSSGP- 1456
Db 1004 AVSQTDSPTTEATPVQEVGGVDP-IEQERRTOEVLOAVAEKVEESQLPCTGQPE 1062
Qy 1457 SVLRSTLH---EAPKAQL-SPGIYDDT-----SARRTPVSYQNTMSRGP-MMNR 1502
Db 1063 DVLPQVQRAEAERPEQAEASGLKKTVDVLKVDQAQAKTEPFTQGVVGTTPSEFEKA 1122
Qy 1503 SDVT--IPPKNKSTHNRKSTILTPQRESIPAKSPVPGVDPVVSHPDPHHRGSTADEVY 1560
Db 1123 PONTESIESELVTTCQAEATLQVKSQEMVNEQAIP---PDSVETPTDSEFDGTPVADF 1179
Qy 1561 WSHLPTQLDPAMPFHRA---LDPA---AAAYLFQRLSPTPGYPQYQLYAMENRTOT 1612
Db 1180 DAPGTTQKDEIVEIHEENEVHLVPVGTAEAVPAQKERPPA---PSSF-VFQETREQS 1235
Qy 1613 ILNDYITSQQMQLRDPDVARGLSPR-----EQPLGLP-YPATRGIIIDLNNPP 1660
Db 1236 KMEDTLEHTDREVS--ETVSILSKTEGTQADQAYADEKTKDVPFFEGLEGSID-TGI- 1290
Qy 1661 TILVPHPGTSTPMDRTYI---PGTQITFPFR---PYNASMSPGHPHTHLAAASA 1712
Db 1291 -----TVGREKVEVALKGEETEAECCKDDALELQSHAKSPSPVEREMVQV 1339
Qy 1713 ERER-ERE---REKERERERIAAASDLYLRPGSEQ-----PGRPGSHGYVRSP 1757
Db 1340 EREKTEAETHVNEKLEHETAVTVSEV-----SKOLLQTVNVPIIDGAKEVSSLEGSP 1394
Qy 1758 SPVKTQETM-----LQORPSVFGTNGTSTVITPLDPTAQLRIMPL-----PAGGPSISQ 1807
Db 1395 PPCLQGEAEVCTKIQVQSSEASFTLTAABEEKVLGETA--NILETGETLEPAGAHVLVE 1452
Qy 1808 GLPASRYNTAA---DALAALVDAASAFQMDYS-----KTKESKHEAAR 1848
Db 1453 EKSEKNEFDAHPGEDAVPTGPPDCAKSTPVIIVSATTKKGLSDDLGEKTTSLKWSDE 1512
Qy 1849 LEENLRSRAVS---EQOOLEQKTLVEKRS-----VOCLYTSSAFPSG 1890
Db 1513 VDEQVACQEVKVSVAIELEPENGILELETKSKLVONITQTAVDQVFRTEETATEMLTS 1572
Qy 1891 KPQPHSSVVYSEAGDKGPPPKSRYEE-ELNRGKTTITANFIDVITTRQIASDKDARE 1949
Db 1573 ELQTOAHVIRADS-QDAQETEKEGEEPQASQAQDETPTS- 1612
Qy 1950 RGSOSSSSSSLSHRVETPSDALEVISPASSPAPPEKLOTQTOPEVVKANQANDPTRO 2009
Db 1613 --KEESESTAVGOAH-----SDISKDMSEAS-----EKTMTVEVE-----GSTVND--QQ 1653

Qy 2010 YEGPLHHRPQESPSQOOLPPSSQAEGMG--QVPRTHRLTLTADHICOLIITO----- 2061
Db 1654 LE-----EVLVPSSEEGGAGTKSVPEDDHALLAERIEKSLVPEPKEDEK 1698
Qy 2062 ----DFARNQVSS-----QTPOQPPTSTFONSPS-----ALVSTPVTRKTS 2098
Db 1699 GDDVDDPENQNSALADTDASGLTKESPDT----NGPKQKEKEDAQVEQLQEGKVHSED 1754
Qy 2099 NRYSPSQASQVHHQRPGRS 2118
Db 1755 KAITPQAEELQKQERESAK 1774

RESULT 20
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montinuy, Marc R.
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; TITLE OF INVENTION: RESPONSIVE GENES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/194,468
; FILING DATE: 10-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9672
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)-546-4737
; TELEFAX: (619)-546-9392
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-194-468-2

Query Match 2.4%; Score 309.5; DB 1; Length 2441;
Best Local Similarity 18.7%; Pred. No. 1.5e-10;
Matches 495; Conservative 304; Mismatches 992; Indels 859; Gaps 122;

Qy 7 PPNQGAFTSQSRYPHPSVQYTFPTNRHQEFVAVPDYRSSHLEYSOASQLQOQOQOQO--- 64
Db 121 PLNQGDST-----PNLPKQA-----ASTSGPTPPASQALNPQAQKQV 159
Qy 65 LRRRPSLLSEFHGSDRPPQRRYSYEFPHG--PSPVDHDSLESKRPLRQVSDSHF--- 119
Db 160 LVTSSTPATSQTGPGIC----MNAFNQTHPCLLNSNSGHSILMNOAQOQOQVWNGSLGAA 215
Qy 120 QRVSAAVPLVLVHPLPEGLRASADAKK---DPAFGGKHEA----- 155
Db 216 GRGAGAGMPYAPAMQAGATSSVLAETLTQVSPQWAG-HAGLNTAAQAGMTKMGTTGTSP 274
Qy 156 ---PSSPISQPCGDDQNASPSKLSKEELIQSDMDRVDREIAKVEQQILKLLKKKQOOLEEE 212
```





|    |      |                                                               |                                               |      |
|----|------|---------------------------------------------------------------|-----------------------------------------------|------|
| Qy | 1254 | IKOGMSNRESVP                                                  | SAPLEGLICRALPRGSPHSD-----LKERTVLSGSIMOGTPRATT | 1305 |
| Dz | 1343 | VNFRLRRQHNEAG-----EVFVRVVASSDKTVEVKPGMKSRFVDGSEMSEFPYRTK      | 1395                                          |      |
| Qy | 1306 | ESFEDGLKYPQOLKRESPP                                           | IRAFEGAITKGPDYGITTIKEMGRSIIHE---IPRODILT      | 1361 |
| Dz | 1396 | ALF-----AFE-----EIDGV-DVCFFGMHVODTALIAPHQ----                 | 1425                                          |      |
| Qy | 1362 | QESRKTPVVOSTRPIIEGS--ISOGTPIKFDN-NSGQSAIKHN-----VKSIL---      | 1407                                          |      |
| Dz | 1426 | -----IQSCVVISYSLDSIHFFRPCLRTAVYHEILIGLEYVKLVYVT               | 1469                                          |      |
| Qy | 1408 | -----TGPSKLSRGMPPLIEIYVENIKVVVERCK-----YEDV--KAGE             | 1442                                          |      |
| Dz | 1470 | AHIWACPPSEGDDYIFHCHPPOOKIPKPKRLQEWYKKMLDKAFAERIINDKYKOIFKQANE | 1529                                          |      |
| Qy | 1443 | TVRSRTSVV-----SSGPSVLKRSTLHEAPKAQLSPGIYDDTSARRTPVSQO----NT    | 1491                                          |      |
| Dz | 1530 | --DRITSAKELYPFGDPFWPNVLESIKEQEBEERKKEESTAETPEGSGDGSKNA        | 1586                                          |      |
| Qy | 1492 | MSRGSPMMNRT-SDVT-----IPPKNSTNHERRKSTLTPTQRES-----IPA          | 1531                                          |      |
| Dz | 1587 | KKNKKTKNNKSSISRANKKPSMNPVSNDSLQKLYATMEKHKEVFVIHLHAGPVIST      | 1646                                          |      |
| Qy | 1532 | KSPVPGVDPVVS-----HSPF-----DPHURGSTAGEVYWS-----                | 1562                                          |      |
| Dz | 1647 | QPPIVDPDLLSCDLMDGRDAFLTLDARKHWEFSLSLRKSKWSTLCMLVELHTQGODRFVY  | 1706                                          |      |
| Qy | 1563 | -----HLPTO-----LDPAMPFHRL-----DPAAAAYL                        | 1586                                          |      |
| Dz | 1707 | TCNECKHHVETRHCIVCEDYDLINCYNYSHTXHKVKNGLGILDDESGSQGEPSQSPQ     | 1766                                          |      |
| Qy | 1587 | FOROLS-----PPOYGPYSOYOLIYM                                    | 1606                                          |      |
| Dz | 1767 | ESRRLSTQRCSIOSLVHACOCRNAGSLPSCOKMRKVVOHTKGCKRKNTNGCPCVKOLIAL  | 1826                                          |      |
| Qy | 1607 | -----ENTROTILNDYTISOQM-----VNLPDPVARGL-SP                     | 1637                                          |      |
| Dz | 1827 | CYTHAKHCENKCPVPFLNIKHNYRQOOIOHCLOQAOLMRRRMATWNTRNVFOOSLPSP    | 1886                                          |      |
| Qy | 1638 | REOPLGLPYATRGIIDLTNMPTTLVPHPGGTSTP-PMDRI-TYIPCTQTITPPP-----   | 1690                                          |      |
| Dz | 1887 | TSAPPGETP-----TQQPSTQTQPPAQOPSPVNNSPAGFNVAQTPTTVISA           | 1936                                          |      |
| Qy | 1691 | -RPNYSASMSGPHTHLAAAASAEREREKERERERIAAASDYLRPGSEQPGRPG         | 1749                                          |      |
| Dz | 1937 | GKPTNOVPAPPPPAOPPPAAVAAARQIEREAQOOOHLYR---ANINMGPMGRDGMGTGP   | 1993                                          |      |
| Qy | 1750 | SHGYVRSFPSVRTQETMLQORPSVFOGTNGTSTVITPLDPTAQLRIMPLPAGGPSISQGL  | 1809                                          |      |
| Dz | 1994 | SQ-----WTPVGLNWPRNQVSGFVWSSM-----PPGQMQQAPIFOQQP--MPGM        | 2036                                          |      |
| Qy | 1810 | PASRYNTAADALAALYDAAASPOMDVSKTESKHEARLEENLRSAVSEOOQLEQK        | 1869                                          |      |
| Dz | 2037 | P-----RVMSMQAOAAVAGPRMPNPQNRSISPSPA-LQDLLRTLKSPSSPOQO--QQ     | 2086                                          |      |
| Qy | 1870 | TLEVEKRSVQCLYTSAF-----PSGKPOP-H-----SSV                       | 1998                                          |      |
| Dz | 2087 | VLNILSNPOLM---AAFKTORTAKYVANOPCMQPPQGLQSQCMPQPGMHQOQPSIQNL    | 2143                                          |      |
| Qy | 1899 | VYEAGKDK-GPPPKSYREEELTRGKTTITAANFID-----VLIITROI-             | 1941                                          |      |
| Dz | 2144 | NAMQAGVPRGVPPOPAMGLNPOGO-----ALNIMNPGHPNMTNMPQYREVMRQQL       | 2199                                          |      |
| Qy | 1942 | --ASDKDAREGSSQSDSSSLSS-----HRYETPSDAIEVISPAS-SPAPPOEKLOTYQ    | 1993                                          |      |
| Dz | 2200 | QHOOOQOQOQOQOQOQOONSASLAGMGACHSFQOQO-----GFGYAPAMQOQRMQOHL    | 2253                                          |      |
| Qy | 1994 | P-----EVVKANO-----AENDPTRQ---YEGPLHYHRQOQSPSPQOQLPP           | 2032                                          |      |
| Dz | 2254 | PIOGSSNGMAAPMGOLGOMQOGLGADSTNIGIQAALQORITLOQOQMKQITGSPGPONPM  | 2313                                          |      |

```

Qy 2033 SSQAEAGQVPRTHRLITLADHI--CQITQDFARNQVSSOTFPQPPPTSTFQNSPSALVS 2090
| | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2314 SPOQHMLSGOQP-----ASHLPGGQIATS--LSNOVRSAPAPVQSPRPQ-----S 2355
| | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 2091 TPVTKTSNRYSPEAQSAQVHH-----QRPGRSVSPENLVKSGRGRPKCS----- 2136
| | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2356 QPHRSSPPIRQQP---PSPHHVSPOTGPHPEGLAVTMASSMDQGHGLGNPEQSAMLPOLN 2412
| | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 2137 -PERSHVS 2145
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2413 TPNESALSSE 2422
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 22
US-08-459-568-4
; Sequence 4, Application US/08459568
; Patent No. 5811304
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,568
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-459-568-4

```

```

Query Match 2.4%; Score 309; DB 2; Length 1719;
Best Local Similarity 18.5%; Pred. No. 9.7e-11;
Matches 344; Conservative 247; Mismatches 664; Indels 600; Gaps 82;

Qy 502 NQOIARPSQEEKVEKEEDKAETKTEKKEEDKEDEKEDSKENTKEKDIQDTAET 561
 ||| :|| :
Db 145 NPEIAAAEEERASARSRSPKSH-----GKKKSEQNKNKNKIQDIQLT 192
 ||| :|| :
Qy 562 EREQATPRGKTANSQGRRRGRITRSMTEAAAAASAAAAAETPPPLPPP-----PE 616
 ||| :|| :
Db 193 SEPDTFSANMRDSABGPKDEDEKSPSALEQPATLIQ-----EVASQEVPPELATPAPAWEPQ 249
 ||| :|| :
Qy 617 PISTEPVTSRWTEEMEYAKGLVEHGRNMAAIAKMVGTTSEAOCKNFYFNKRRHMLD 676
 ||| :|| :
Db 250 P-----RP-----DRLREA-----AAEE-----AAEE-----VN 264
 ||| :|| :

```





SEQUENCE CHARACTERISTICS:  
LENGTH: 1719 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-399-411-4

Query Match 2.4%; Score 309; DB 2; Length 1719;  
Best Local Similarity 18.5%; Pred. No. 9, 7e-11;  
Matches 344; Conservative 247; Mismatches 664; Indels 600; Gaps 82;

QY 502 NQOIARPOEKVEKEEDKAEKTEKKEEKKDEEKEDEKSKTEKIDKIDTABEL 561  
DB 145 NPEIAAIEERASAKRSRSPKSRK-----GKKQENKNGNKIQIDQLKT 192

QY 562 EREQATPRGKTANSQRRKGRITRMTNEAASAAAAAATAEPPPPPLPP-----PE 616  
DB 193 SEPDTSANMRDSABGKDEKPSASALEOPATLQ---EVASQEVPPPELATPAPAWEPQ 249

QY 617 PISTEPVETSRWTEEMEVAKKGLVHGCRNAAIAKVMGKSEAOCKNFYFNYKRRHND 676  
DB 250 P---EP-----DRLEA-----AAE-----VN 264

QY 677 NLLQHKQKTSKPREERDVSOCEASVASTVSAQEDIEDAENNEENPEDSEYAVKPS-- 734  
DB 265 DLGESEEEDEEED-----DDDELEDEGESEASMPNENSVKEPEIR 311

QY 735 -EDSPENATSRGNTPAVELEPTTETAPSTSPSLAVPSTKKAEDSVETQVNDISAE 793  
DB 312 CEKPEDLLE---EPKTTSETLEDCEVTPAMQIPRTK----- 347

QY 794 EQMDVQOEHSAEGSCDPPPPATKADSVYDEVRY---PENHASKVEGDNTKERDLDRAS 850  
DB 348 EBANGDVETFMPCQCHERETTK-QGLERHMHITVNHAFKCKYCGK----- 397

QY 851 EKVEPRDELDVVAQOINAQRPQSD-----NDSSATCSADEVD-DCEPERQMFPM-DSK 904  
DB 398 -----AFGTQINRRRRHRRHEAGLKRKPSQTLQPSDELADKAGSENVASKDDSS 447

QY 905 PSLNPTGSILVSSPLKPNLDLPOLQHRAAVIPWVSTPCNIPICGTPVSGYALYQRIH 964  
DB 448 PPSLGPDLIMNSEKASQDTINSVVEENGEV-KELHCKYCKKVFGT-----HTNMRHQ 502

QY 965 KAMHESAL-----LEORORQEQIDLECRSTSPGCTSKPNREWEVLQAPQOL 1014  
DB 503 RVHERHLIPKVRKKGLEBPQPAEQ-----AQATQNVYVPSTEPGEAGEADVVIMDI 558

QY 1015 ITNLEGV-----RLPTTRTPRPPPLIPSKTTVASEKPSFIMGSSISQGTPTGLTS 1068  
DB 559 SSNISENLNYIDGKIQTNNNT-----SNCDVIEMESASADLYGINCLLTPVTVEIT 610

QY 1069 HN-----QASYTQETPKPSVGSISGLPQOBSAKSATLPIYKQEFSS-----PRSONSOPE 1120  
DB 611 QNIKTQVPTVEDLPELGTSTN-SEAKKRRTASPPALPKKAETDSDPMVPSCLSPL 669

QY 1121 GL-----LVRAQHE-----GVVRGTAGAIQIE-GSTRTGTPTSK 1152  
DB 670 SISTTEAVSFHKEKSVYLSSKLQLQLOQDKLTTPAGI---SATEIAKLGPCVCSAPASM 726

QY 1153 ISVESI-----PSLRGSITQGTTPALPOTGIP-----TEALVKGISIRMPEDSSPE-- 1198  
DB 727 LPVTSSRFKRRITSSPPSPQHSPPALRDFGKPSDGAAMTDAGLTSKKSKLESKSHSDSPA 786

QY 1199 -KGREAAASKHVIYE-----GKSHILSYDN-----IKNAREGTRSPRTAHE 1240  
DB 787 LSGRDERETVSPPCDEYKMSKEWTASSAFSSVCNQOPLDLSSGVKQKAEGTGKTPVQWE 846

QY 1241 ISLKPSYESVEGNIQKQMSMRSPVSALEGLICRALPRGSPHSLDKERTVLSGSIMOGT 1300  
DB 847 SVLDLSVHKHCKSCSEGEKESHVQBP-----TCSAVKRRKPTTCMLQKVLNNE----- 896

QY 1301 PRATTESPDGLKYPKQIKRESPPPIRAFEAGITKGPYDGIITIKEMGRSITHEIPRODIL 1360

DB 897 -----YNGIDLPE-----NPADG-TRSPSPCKSLEAQDPDPLG 929

QY 1361 TQESRKTEVQVOSTRPIIEGSIQGTPIKFDNNSQSAIKHNKSLIITGPSKLSRGMPL 1420

DB 930 PGSGPAP-TVESTPDVCPSPALQTP---SLSSQO-----LPLP 965

QY 1421 EIVPENIKWVERGKYEDVKAGETVRSRHTSVSSPSVLRSTLHEAPKAQLSPIYDDTS 1480

DB 966 -LIPD-----PSPPPCPVLTATPPPLPTVPLPAPSSASP----- 1005

QY 1481 ARRTVSYQNTMSRSGPMNRTSDVTIPPNKSTNHERKSTLTPQTQRESIPAKSPVGVDP 1540

DB 1006 -HPCP-----SPLSNATAQSPLP-----ILSPTVS---PSPSPIPVPEP 1040

QY 1541 VVS-HSPFDPHRRGSTAGEVYWSHLPTQDLPAMPFHRALDPAAAYLQFQRLSPGTPPS 1599

DB 1041 LMSAASPGPPLTSSSSSSSSSSSSSSSSSSSSP-----SPPP----- 1078

QY 1600 QYQLYAMENTRQITLNDYITSOQMOVNLRPDVA---RGLSPRQPLGL----- 1644

DB 1079 ---LSAISVYSS---GDNLEASLPMISFKQELENEGLKPREEPOSAAEQDVVQVETNK 1133

QY 1645 -----PYPATRGII-----DLJNMPPTILVPHPGGT--- 1670

DB 1134 NFVCNVCESPFLSKIDLTKHLSIAHEWPFKCEFCVQLFKDKDLSEHFLHLGVGNIFV 1193

QY 1671 -----STPPMDRI---TYIPGTQITFPFPPRYNSASMSGPHTHLA 1707

DB 1194 CSVCKKEFAFCLNQOQDRLHP-DKVCTHHEFEGTL-----RPNFTDPSKAHVEHQ 1247

QY 1708 AAASAEERERERERERERERERERERERERERERERERERERERERERERERER 1767

DB 1248 SLP-----EOPLETSEEEELNDSSEELYTIKIMASG-----IKTKDPRVLG--- 1291

QY 1768 LQQ-----RPSVFQ-----GNGTSVITPLDPTAQLRIMPLPAGGPSISQGLP 1810

DB 1292 LNHQVSEFKPPFPFYHHRNPMGIGVTATNFTTNIPQTFTTATIRCTKCGKG---VDNMP 1347

QY 1811 ASRYNTAADALAALVDAASAPO---MDVSKTKESH--- 1844

DB 1348 ELHKHILACASAS---DKKRYTPKKNPVLKQTVQPKGVVLDNSGKNAPFRMGQPKRLN 1405

QY 1845 ---EAARLEENLRSSAAVSEQQLEQKLEVEKRSVQ-----CLYTSSA 1886

DB 1406 FVELSKMSSN-KLKLNALKKQLVQKAILQKNKSAQKADLNACESSHICPYCNRE 1464

QY 1887 FP-SGKPPQHSVYSEAGKDKGPPPKSRYEELRTRGTTITTAANFIDVITROIADSK 1945

DB 1465 FTYIGSLNKHAA--FSCPCKPLSPPKK-----KVS 1492

QY 1946 DARERGSOSSSSS---SLSSHRYETPSDAIEVISPASSAPPQKLOTQYQEVVKANQA 2002

DB 1493 HSSKKGHSSPASSDKNSNHRRET-----ADAEIKMSMOTPLGKTRAR 1538

QY 2003 ENDPTQREQGLHRYRPOE---SPSQOQLPPSSQAGMGQVPRTHRLITLADHICQII 2059

DB 1539 SSGPT-QVPLDSSSRFQKNYKFAASVKSPSSSLNNSPI-----RWAKIT-HVEGKK 1592

QY 2060 TQDFARN---QVSSQTQPPPTSTFQNSPSAL--VSTPVRTKTSNRYSPESQAQS 2109

DB 1593 PKAVAKNHAQLSSKTSRSLHVRV-QSKAVLQSKSTLASKKRTDRENIKRSRS 1646

RESULT 24  
US-08-516-859A-4  
; Sequence 4, Application US/08516859A  
; Patent No. 6069231  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 106



Qy 1845 ---EAARLEENLRSSAAVSEQQLEQKTELEVKRSVQ-----CLYTSSA 1886  
Db 1406 FSVLSKMSNN-KLKLNALKKKQLVQKAILQKNKSAKADLKNACSSSHICPNCRE 1464  
Qy 1887 FP-SGKPPQPHSSVYSEAGKDKGPPPKSRYEELRTRGKTTITANFDIVITRIASDK 1945  
Db 1465 FTYIGSLNKHAA--FSCPCKPLSPPK-----RVS 1492  
Qy 1946 DARERGSQSSSS---SLSSHRYETPSDAIEVISPASSPAPQEKLOTYOPEVVKANQA 2002  
Db 1493 HSSKKGHSSPASSDKNSNHRRT-----ADAEIKMSMQTPLGKTRAR 1538  
Qy 2003 ENDPTRQVEGLHHYRPOE---SPSPQQLPPPSQAGMGQVPRTHRLITLADHICQII 2059  
Db 1539 SSGPT-QVPLPSSFRSQNVKFAASVSKKPSSSSLRNSPI-----RWAKIT-HVEGK 1592  
Qy 2060 TODFARN---QVSSQTPQOPTISIFONSAL--VSTPVRTKTSNRYSPESQAOS 2109  
Db 1593 PRAVAKNHSQLSKTSLSLHVRV-QSKAVLQSKSLASKKRTDRFNKRSRS 1646

RESULT 25

US-08-046-585-5  
; Sequence 5, Application US/08046585  
; Patent No. 5453362  
; GENERAL INFORMATION:  
; APPLICANT: Lamarco, Kelly  
; APPLICANT: Wilson, Angus  
; APPLICANT: Hertz, Winship  
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:  
; TITLE OF INVENTION: HOST CELL FACTOR  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/046,585  
; FILING DATE: 12-APR-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36, 627  
; REFERENCE/DOCKET NUMBER: A-57503-1/RAO  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2035 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-046-585-5  
  
Query Match 2.3%; Score 294.5; DB 1; Length 2035;  
Best Local Similarity 19.5%; Pred. No. 9.8e-10;  
Matches 412; Conservative 237; Mismatches 839; Indels 623; Gaps 98;  
  
Qy 540 EKESKNTKEKDKIDGTAET---EEREQATPRGRK-----TANSQGRRK 583  
Db 287 EKWKCTNTLACLDNMTAWETILMDTLIEDNIPRARAGHCAVAINTRLYINSGRGRYKA 346

Qy 584 RITRSMTEAAAAAASAAAAATEEPPPLPPPEPISTPVPVETSRWTEEMEVAKGLVEH 643  
Db 347 -----WNNQVCCCKDLWLETEKPPP-----ARVQLVRANTNSLEV----- 382  
Qy 644 GRNNAIAIAKAVGTGKSEACKNFYFNKRRHNDLNLQHQKOTSKRPREERDVSCESA 703  
Db 383 ---SWGAVA-----TADSVLLQ-----LQKYDI---PATA 406  
Qy 704 STVSAQEDIEDIEASNEENPEDSEAVKPSDESPENATS--RGNTEPAVELEPTEAP 761  
Db 407 AT-----ATSPTPNPVPS--VPANPKSPAPAAAAPAVQPLTQVGTITLLPQAAP 455  
Qy 762 STPSLAYPSTKPADESQVNDISAEATAEQMDVQDQEHSAEEG-----SVCDDPPA 816  
Db 456 PTTITIQVLPIVP-----GSSISVPYA-----ARTQGVPAVLKVTGPQAT 495  
Qy 817 TKASVDVEVRVPEHNHASKVEGDNTKERDLORASEKVEPRDEDLVAAQ-----INAORP 871  
Db 496 TGTPLVTMR---PASQAGKAPVTVTS-----LPAGVRMVVPTQSAQGTIGSSP 541  
Qy 872 EPQSDNDSSATCSADEVDGEPERQMFMDSKPSLLN-PTGSILVSS-PLKPNPLDLPQ 929  
Db 542 QMSGMAALAAAAATQKI-----PPSAPTPLVSPAGTTIIVKTMVTPGTTTLLPA 591  
Qy 930 LQHRAAVIPPMVSVCTPCN-----IPIGTPVSGYA-LYQRHIKAMHESALLEEQRQOE 981  
Db 592 TV-KVASSPVMVS-NPATRMLKTAQAQVGTSSVSSATNTSTRPIITVHKSGITVVAQQAQ- 648  
Qy 982 QIDLECRSTSPCGTSKSPNREWEVLQAPAHQIINLPEGVRLPTRP----- 1029  
Db 649 VVTVVGGVTKTITLVKSP-----ISVPGGSAISNLKVMVSVQTKPVQTSVAVTQAST 703  
Qy 1030 -----TRPPPLIPSSK-TTVASEKPSFIMGSIQGTPTGLTSHNQASYTQETPK 1080  
Db 704 GPVTQIIQTKGLPAGTILKLVTSADGKPTTIITQTQASGA-GTKPTILGISSVSPSTK 762  
Qy 1081 PSVGSISLGLPRQ-----QESAKSAT-LPYIKOEESP-----RSONSOPEGLL--- 1123  
Db 763 PCTIITIKTIPMSAIIIOAGATGVTSSPGIK---SPITIITKVTMTSGTGAPAKIITAV 818  
Qy 1124 ---VRAQH-----EGVVRGTAGAIQEGSITRGT-----TSKISVESI-PSLRGSIQOG 1168  
Db 819 PKIATGHGQGGVTVQVLKGAQP--QPGTILRTVPMGVRLVTPVTVSAVKPAVTTLVVKG 876  
Qy 1169 TPALPQTCIPIEALVKGSIRMPIDESSPEKREAAKSHVIEGKSHILSYDNINA 1228  
Db 877 T-----TGVTTLGTVTGTVS-----TSLAGAGHSTSSASLATPITLTGIATL 919  
Qy 1229 REGTRSPRTAHEISLRSYVESVEGNIKQGMSPVSPAPLEGLICRALPRG---SPHSD 1285  
Db 920 SSOVINP-TAITVSAAOITLTAAGGLTT-PTITMQVPSQPTQVTLITA-PSGVEAQPVHD 976  
Qy 1286 LKERTVLGSGSIMQGTPTATTESFEDGLKYPKOIKR--ESPPIRAPEGAITKGPYDGTIT 1343  
Db 977 LPVSLASPTTEQPTATVTIADSGQGVQPGVTVLVCSNPPCETHETGTNTAT---TTV 1033  
Qy 1344 IKEMGRSIEHPIRQ-DILTQESRKTPEVWQSTRPIIEGSIQ--GTPIKFDNNSQSAIK 1400  
Db 1034 VANLGG---HPQPTQVQFVCDRQEAASLVTSTVGQNGSVVVRVCSNPPCETHETGTNTA 1091  
Qy 1401 HNVKSLITGPKSLRGMPPLEIVPENIKVVERKGYEDVKAGETVRSRHTSVSSGPSVL 1460  
Db 1092 TTATSNMAGHGCNSN--PPCE-----THET----- 1114  
Qy 1461 STLHEAPKAQLSPGIYDTSARRTPVSVQNTMSRSPMMNRSDVT-----IPPKNKSNH 1515  
Db 1115 GTTNTATTAMSSVGNHQRDARR-----ACAACTPAVIRISVATGALEAAQGSQCK 1167  
Qy 1516 ERKSTLPTQRESPAKSPGVCDPVVSHSPFDPHHRGSTRAGEVYVWSHLPTQLDPMPFH 1575  
Db 1168 TRQTSATSTTMVATGAP-----CSAGPL-----LGPSM--- 1197



|    |      |                                                               |      |
|----|------|---------------------------------------------------------------|------|
| Qy | 930  | LQHRAAIVPMWVSCPCN-----IPGTGPSGYA-LYQRHIAMIESALLEQRORQE        | 981  |
| Dz | 592  | TV-KVASPPWVS-NPATRLKLTAAAOQGVTSVSSATNTSTRPIITVKHSGTVTVAOQAQ-- | 648  |
| Qy | 982  | QIDLECRSSTPCGTSKSRENWEVLQAPHOIINLPEGVRLPTTRP-----             | 1029 |
| Dz | 649  | VVTVVGGVTKTILVKSP-----ISVFGGSALISNLKVMXSVQTKPVQTSAVTGQAST     | 703  |
| Qy | 1030 | -----TRPPPPLIPSK-TTVASEKPSFIMGSIQSQTPTGLYLASHNOASYTQETPK      | 1080 |
| Dz | 704  | GPVTOIIHQKGPLAGCTILKLVTSDAGKPTTIITTQASGA-GTKPTILIGISSVPSTTK   | 762  |
| Qy | 1081 | PSVGSISSLGRQ-----QBSAKSAT-LPYIKOEERFP-----RSQNQQPELL---       | 1123 |
| Dz | 763  | PGVTTIKITPMSAIITQAGATGVTSSPGIK---SPITIITTKVMTSGTGAPAKIITAV    | 818  |
| Qy | 1124 | --VRAOH-----EGVRGTAGAIOEGSITRGRP-----TSKLSVESI-PSLRGSTQG      | 1168 |
| Dz | 819  | PKIATHGGOQVTVQVWKAGP--QPOTILRTVPMGVRLVTPVTVSAVKPAVTVLVKG      | 876  |
| Qy | 1169 | TPALPOTGITPEALKRGISRMPIEDSSPEKGREEAASKGHVIEKSGHTLSVDNIKNA     | 1228 |
| Dz | 877  | T-----TGVTLGTVTGTVS-----TSLAGAGHSTSASLATPTITLGTATL            | 919  |
| Qy | 1229 | REGTRSPRAHEISLKRSYESVEGNIKOMSNRESVPVSAPLEGICRALPRG----SPSHD   | 1285 |
| Dz | 920  | SSQVIN-TAITVSAOQTTLAAGLLT-PTITMQPVSOPTOVTLITA-PSGVEAQPVHD     | 976  |
| Qy | 1286 | LKERTVLGSIQGTPRATESFEOLGLYPKOIKR--ESPPIFAEGAIHKRPDYGIITT      | 1343 |
| Dz | 977  | LPVSIILASPTEOPTATVIADSGOGDVQPGCTVLVCNSNPCETHETGTTNTAT---      | 1033 |
| Qy | 1344 | IKEMGRSIEHIPRO-DILTQESKRKPEVVQSTRPIEIGISIQ--GTPIKEDNNSGQSAIK  | 1400 |
| Dz | 1034 | VANLGG--HPQPTQVOFCVCDROEAAASLVTSTVGQONGSVVRCNSNPCEHETGTTNTA   | 1091 |
| Qy | 1401 | HNVKSLITGSKLSRGMPPLEIVPENIKVYERKGYEDYKAGETVRSRHTSVSSGSPSVLR   | 1460 |
| Dz | 1092 | TTATSNMAGHGCSN--PPCE-----THET-----                            | 1114 |
| Qy | 1461 | STLHEAPKAQLSPGIYDDTISABRTPVSYONTMSRCSPMMNTSDVT-----IPNKSTNH   | 1515 |
| Dz | 1115 | GTTNTATTAMSSVGANHQDRRR-----ACAAGTPAIVIRISVATGALEAAOGSKSOC     | 1167 |
| Qy | 1516 | ERKSTLTPTQRESIPAKSPVPGDPVWSHSPDPHHRGSTAGEVYVSHLPTQLDPAMPFH    | 1575 |
| Dz | 1168 | TROTSATSITWTMATGAP-----CSAGPL-----LGPSM---                    | 1197 |
| Qy | 1576 | RALDAAAAYLFQROLSP-----TPG--YPSQYQLYAMENTROTILNDYITSOOMQ       | 1624 |
| Dz | 1198 | -AREPGGRSPAIFY-QLAPLSSKVLRLSSPSIKDLPAGRHSNAVSTAAMT-----       | 1243 |
| Qy | 1625 | VNLRPDVARCLSPREOPL--GLPYATRGIDLTNMPTTILVPHPGGT--STPPMDRIT     | 1679 |
| Dz | 1244 | ---RSSVYGAG-EPRMAPVCESLOGGSPSTTVTVTAL-BALLCPSATVTCVSNPPCB--T  | 1296 |
| Qy | 1680 | YIPGTQITPPRPYNASMSPGHP-----THLAAAAA-----EREREREKE             | 1723 |
| Dz | 1297 | HEYGTTNTATTSNAGSAQVRCNSNPCEHETGTHTATNTATNSNGTGOPEGQQPPAGR     | 1356 |
| Qy | 1724 | RERERIAASSDLYLRPGSEQPGRGSHGYVRS-----PSPSVRTOETMLQORPSVFQGT    | 1778 |
| Dz | 1357 | CETHQTTSTGTTMSVSVGALLPDATSSHRTVESGLEVAAPSVTPQA-----           | 1403 |
| Qy | 1779 | NGTSVITPLDPAQLRI MPLPAGGFSISQGLPASRYNTAADALAALYDAASAAPQMDSVK  | 1838 |
| Dz | 1404 | -GTALLAPP-PQORVCNSP-----PCETHETGTTHTTATTVTSMNSSNQ-DPPP        | 1448 |
| Qy | 1839 | TKESKHEARLEENLRSAAYSEOOOLEOKTLEVEKESVOCLYTSSTAFPSGRQPQHVSV    | 1898 |
| Dz | 1449 | AASQGEV-----ESTQGDENVNITSSTATTVTSTLTRAFTVTQSTVPVGPSPVPPPEEL   | 1504 |
| Qy | 1899 | VYSEAGDKGPPPKSRYEEBLTRGKTII--TAANFDIVITROIADSKDAREGROSSD      | 1956 |

RESULT 27  
PCT-US93-11721-5  
Sequence 5, Application PC/TUS9311721  
GENERAL INFORMATION:  
APPLICANT: Lamarco, Kelly  
APPLICANT: Wilson, Angus  
APPLICANT: Herr, Winship  
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:  
TITLE OF INVENTION: HOST CELL FACTOR  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/11721  
FILING DATE: 03-DEC-1993  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A.  
REGISTRATION NUMBER: 36,627





QY 1118 QPEGLVRAQHEGVVRGTAGAIQBSITRGTPTSKISVESIPSLRGSITOGTPALPOTGI 1177  
DB 995 QQE---ERGQOE---REVARLTQE---RGRAQADLALEK---AARAEI-----1030  
QY 1178 PTEALVKGISIRMPIDESPSPKREAAASKGHVITYEKGSHILSYDNKIKNAREGTRSPRT 1237  
DB 1031 --EMRLQNALNEQREFAT-----LQEAHALATEKEGD-----1063  
QY 1238 AHEISLRSVESVGNIKQGHSMRESVSPAPLEGICRALPRGSPHSDLKERTVLGSSIM 1297  
DB 1064 -QELAKLUGLEAAQ--IKLEELROT-VKQLKEQLA-----KKEKEHASGGA 1107  
QY 1298 QCTPRATTESPEDGLKYPKQIKRESPPITRAPEGAITKGKPYDITTIKMGERSITHEIPRO 1357  
DB 1108 QSEAGRTEP-----TGPKLEALRAEVSK-----1131  
QY 1358 DILQESRKTEPVVQSTRIPIEGSISOGTPIKFDNNSGQSAIKHN--KSLITGPSKLSR 1415  
DB 1132 --LEQOCQKQEQADS-----LERSLEAERASRAERDSALETLQGLEKEAQELGHQSAL 1185  
QY 1416 GMPLEIVPENIKVVERGKYED-----VKAGETVRSRHTSVVSSGSPVLKSTLHEAPKAQ 1470  
DB 1186 ASAQRELAFAFTKYVDHSHKSAEDENKAQVARGRQBAERKNSLISS-----1229  
QY 1471 LSPGIYDDTSARRTPSVYONTMGRSGPMNRTSDVTIPPNKSTNHERKSTILTPTQRESIP 1530  
DB 1230 ---LEEVSS-----ILNR-----QVLEKEGESKELKRLVMAESEKS--1262  
QY 1531 AKSPVGVDPVVSHPDPHRRGSTAGEVYWSHLPTQLDPAMPFHRALDPAA-----AA 1584  
DB 1263 -----QKLESCACCRQROPATVPDELQNA 1287  
QY 1585 YLFORQLSPTPGYPSQYQVLYAMENRTQILNDYITSQOMQVNLDPDVARGLSPREQPLGL 1644  
DB 1288 LLCGRRCRAS--GREAEKORVASENLROBELTSQAERAEELQELKAWQEK--FFQKEQALS-1344  
QY 1645 PYPATRGILDLTNMPPITLVPHPGGTSTPPMDRITYIPGTITPTPPYNSASMSGHPT 1704  
DB 1345 -----TLQLEH---TSTQAL-----VSELLPA--K 1364  
QY 1705 HL-----AAAASREREREREKERE-----RERIAAASDL-----YLRGSEQPGRGS 1750  
DB 1365 HLCQQLAQEAQAAAEKREELQESQKAGGLRAELLRAELQAGELGELILPRLQKVAEQDRTAQ 1424  
QY 1751 HGYVRSFSPSVRTOETMLQORPSVFQGTNGTSVITPLDPTAQLRIMPLPAGGPPSISGLP 1810  
DB 1425 Q--LRAEKASTAEQLSMLKKAHGLIAEEN-----RGL-1454  
QY 1811 ASRYNTAADALAAALVDAASAPQMDVSKTKESKHEAARLEENLRSRAAVSEQOQLBQKT 1870  
DB 1455 GERANLGRQFLEVELDQA-----REKYVQELAAVRADATRLAEVQREASTARE 1504  
QY 1871 LEVEKRSVQCLYTSSAPPSPGKQPHSSVYVSEAGKDGPPPKPSYBEELTRGKTTTAA 1930  
DB 1505 LEV-----MTAKYEGAKVVLVEERQFOBE---RQKLTAQVE 1538  
QY 1931 NFIDVITRQIASDKDARE-----RGQSQSDSSSLSSHRYETPDSAEIVSPASSPAP 1984  
DB 1539 E-----LSKKLADSDQAKVQOQKLVQAGGESQEQEORF-----1575  
QY 1985 PQEKLOTYQPEVVKANAENDPTROYEGPLHHY--RPOQESPSPOQQLPPSSQ-----2035  
DB 1576 -QAQLNELQALSQKEQAEEHYKLQMEKAKTHYDAKKQONQELQELQSLRLEQKQENKEL 1634  
QY 2036 ---NEGMGQVPRTHRLTL-ADHICQIITQDF--ARNQVSSQTPQOQPTSTFFQNSPSALV 2089  
DB 1635 RAEAEIRGHELQOQAGLTKAEQTCRHLTAQVRSLEAQVAHADQQLRDLGKQFQVATDALK 1694  
QY 2090 STPVRTKTSNRSPESQAQSVHHQRPGRSVSPENLVDKSRGSRGKSPERSHSSSEPEP 2149  
DB 1695 SREPOARPOLDLSDLSLSCBETPLSITS-----KLPRTOQDGTSPVGEPAASP 1744

QY 2150 IS-----PPQV-----PVVHEKQ-----DSL-----LLLSQRGAEPABORN-----2180  
DB 1745 ISORLPKVESLESLYTTPPARSOQAPLESLSLSDGVFLDSGRKTSARRRTTQIINIT 1804  
QY 2181 -----DARSPGSGISYLSFTT-----KLENTSPMVYKSKQEIFRKLNSGGGSDMAAAQP 2231  
DB 1805 MTKKLDVDEEPSAN--SSFYSTRSAPASQASLRATSTQSLAR--LGSPDYCNSALLSL-P 1860  
QY 2232 GTEIFNLPAVTTSCSVSSRG--HSFADPASNLGLEDIIRKALMGSFDDKVE-----DHGVV 2285  
DB 1861 GYR-----PTTRSARRSOAGVSSGAPPGRN-----SFYMGTCQDEPEQLDDWNRIAE 1908  
QY 2286 MSOPMGVVPCTANT-----SVVTSGETRREEDPSPHSGVCKPKLISKSNSRKS 2335  
DB 1909 LQQRNRVCPHLKTCYPLESRPSLSLGTITDEEMKTGDP-----QETLRAS 1955  
QY 2336 KSPI-----PGOGYLGTERPSSVSVHSEGDYHROTPGWANEDRPSSTGTO--FPYNPLTMR 2391  
DB 1956 MQPIQIAEGTGITTRQQRKRVSLEPHQG-----PG-----TPESKKATSCFP-RPMTPR 2003  
QY 2392 -----MLSSTPTPIACAPSAVNOAAPHQ 2415  
DB 2004 DRHEGRKQSTEAQKKAAPASTKQADRRQ 2032  
RESULT 29  
PCT-US93-06160-4  
; Sequence 4, Application PC/YUS9306160  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA HURWITZ & THIBEAULT  
; STREET: 53 STATE STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/06160  
; FILING DATE: 19930621  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/248-7000  
; TELEFAX: 617/248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US93-06160-4

Query Match 2.3%; Score 286.5; DB 5; Length 2101;  
Best Local Similarity 18.4%; Pred. No. 3.2e-09;  
Matches 480; Conservative 336; Mismatches 924; Indels 869; Gaps 110;  
QY 52 QASQLLQOQQOQQOOL-----RRRPS-----LLS--EFPHGSDRPPQRRRTSYEPHPG 95  
DB 48 EQOQLKQPVSERLDFVCSFLQNRKHPSPSECLVSAQKVLGESELELAKWTMLLY---104





Db 1695 SREPAKPOLDSIDSLDSCCEGTPLSITS-----KLPRTOPDGTVPGEPA SP 1744  
QY 2150 IS---PQV-----PVVHEKQ-----DSL-----LLLSQGAPEAEOBN----- 2180  
Db 1745 ISQRLPKVESLESYFTPIPARSQAPLESLSLGDVFLDSGRKTRSAARRTTQIINIT 1804  
QY 2181 -----DARSPGISYLPSTFT-----KLENTSPMKVSKKQETFRKLNSGGGSDMAAAQ 2231  
Db 1805 MTKKLDVEEPSAN--SSFYTRAPASQASLRATSTSTQSLAR--LGSFDYGNLSALLSL-P 1860  
QY 2232 GTEIFNLPVATTSVSVSRG--HSPADPASNLGLIEDIIRKALMGDFDDKVE-----DHGV 2285  
Db 1861 GYR-----PTTRSSARRSQAGVSSGAPPGRN-----SFTMGTCQDEPEGLDDWNRIAE 1908  
QY 2286 MSQPMGVVPGTANT-----SVTSGTTRREEGDPPSHSGVCKPKLISKSNRKS 2335  
Db 1909 LOQRNRVCPPLHKTCYPLESRPSLSLGTITDEEMKTGDP-----QETLRAS 1955  
QY 2336 KSPI---PGQGYLGTERTSPSSVSHSEGDYHRQTPGNWEDRPSSTGSTQ--FPYNPLTMR 2391  
Db 1956 MQPIQIAEGTITTRQQRKRVSLPHOG-----PG-----TPESKATSCFP-RPMTPR 2003  
QY 2392 -----MLSSPTPTPIACAPSVAQAAPHQ 2415  
Db 2004 DRHEGRKQSTTEAQKAAPASTKQADRRQ 2032

## RESULT 30

US-08-466-390-4  
; Sequence 4, Application US/08466390  
; Patent No. 5686562  
; GENERAL INFORMATION:  
; APPLICANT: TOUTATLY, GARY  
; APPLICANT: LIDGARD, GRAHAM P  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,390  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-466-390-4

Query Match 2.3%; Score 285.5; DB 1; Length 2101;  
Best Local Similarity 18.4%; Pred. No. 3.7e-09;  
Matches 480; Conservative 336; Mismatches 924; Indels 869; Caps 110;

QY 52 QASOLLOOQOQOOL-----RRRPS-----LLS--EFHGPSRQDQERTSYEPHPG 95  
Db 48 EQOQILKQPVSRIDDFVCSFLQKNRKHPSPECLVSAQKVLGSELELAKMTMLLY--- 104  
QY 96 PSPVDHDSLESKRPLRQVSDSHFQVSAVPLVPLVHPLPEGLRASAD----- 142  
Db 105 -----HSTMSKSPDWEQEFYKIQAEALAVILKFLVDHEDGLNLEDFLOKAPVST 159  
QY 143 -----AKDPAF-----GKHAPSPISGQCGD----- 167  
Db 160 CSSTFPPEELSPSHQAKREIRFLELQKVASSSSGNFSLGSPAS--PMGDITLQPFQMR 217  
QY 168 -----DONASPSKLSKE--ELIQSDMDVDRETAKEVQOQILKLLKQKQOOLEEAAKPE 218  
Db 218 RLKQLADESRNDELELELAENRKLITEKDAQIAMQQRIDRLALLN--EKQASPLE 274  
QY 219 PEKVPSPVPEQKHSRSTVQIYYDENRKKAAEAAHKIFEGLGPKVELPLYNQPSTKVYHEN 278  
Db 275 PKE-----LEELRDKNESILMRLHETLK-----OCODLKTEKSQ 308  
QY 279 I--KTQVNRKLLIFTKRRNHARKQRE-----QKICQRYDQLMKAEWKKVDRIENPRR 331  
Db 309 MDRKINQLSEENGDLSEKLRFAHLQLOQDALNELTEHSHKATQEWLEKQAKLEKLSA 368  
QY 332 KAKESKTRVYEEKQPEIRKQREOQOERFQ--RVGORGAGLSATIAARSEHEISEIIDGLSEQ 390  
Db 369 ALQDKKCLE-----EKNEILQKLSLEHLSQLQDNPPQEKGEVLQVLQ 415  
QY 391 ENNEKQMRQLSV-----IPPMFDAQRRVKFINNMGLMEDPMKVYKDRQPMNVWTD 442  
Db 416 ETLKQEAATLAANNNTQLQARVEMLETERGOQEAQLLAERGHFEE-----EKQQLSLITD 470  
QY 443 HEKEI-----FKQFIOHPNFGI-----IASYLERKSVPCVLYIYLTKNENYKALVR 492  
Db 471 LQSSISNLQAKKELEQASQAHGRLTAQVAS-----LTSELTTFNATQ 515  
QY 493 RNYGKRGRNQIARPSQEEKVEKEEDKAEKTEKKEE-----KKDEEE 537  
Db 516 Q-----QOELAGLQQAQ--EKQAQLAQTLQOQQAQSOGLRHQVQELSLSLKQEQ 566  
QY 538 KDEKEDSKENTKEDKIDGTAETETEERQATPRGRKTYANSQGRKGRITRSMWNEAAAS 597  
Db 567 LKEVAEQEATRQ--DHAQQLATAAEERE--ASLRERDAALQK-----LEALE 610  
QY 598 AAAAAATEEPPPLPPPEPISTPEVTSRWTEEMEVAKKGLVEHGRNMAAIAKWGVTK 657  
Db 611 KEKRAKLEILQOQLOQVANEARDSAQTSVTAQREKALSRLK--VEELOACVETARQOHE 668  
QY 658 SEAOCKNFYFNKRRHNLNLLQOHKQKTSRKPR--EERDVQ-----CESVASTVSAQ 709  
Db 669 AQAQVA-----ELEQLRSEQQKATEKERVQAEKQLOEQLOALKESLKVTKGSL 718  
QY 710 EDEIEASN--EENPEDESEVAKPS-----EDSPENATSRGNTPEVELEPFT 757  
Db 719 EEEKRAADALEEQORCISELKAETRSVLVQHKRERKELEFEERAGRKGLEARLLQ----- 774  
QY 758 ETAPSTPSLAVPSTKPAEDESVEVQVNDSTSAETAQMDVDQOEHSAAEGSVCDPPAT 817  
Db 775 -----GEAQAQATEV---LRRELAEM---AAQHTAES----- 801  
QY 818 KADSVQVVEVRVPHNSHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQOQINNAORPEPSDN 877  
Db 802 -----ECEQLVKEVAWR---DGYEDSQEBAQYGAMFQBLMTLKECEKARQELQ--- 850  
QY 878 DSSATCSADEVDGEPERQRMFMDSKPSLLNPTGSLVSSPLKPNLDLPOLQHRAVI 937  
Db 851 -----EAKKQVAG-----IESH-----SELQISROQNKLAEL 877  
QY 938 PMVSCFPCNIPGTIPVSGVGYLQYORHIKAMHESALLBEORORQOIDECKRSSSPGCTS 997  
Db 878 HANLA-----RALQVQEKVEYRAQKLAADDLSTLQEKM-----AATS 913  
QY 998 KSPNREWEVLQAPAHOLITNLPBGVRLPTTRTPPPPLIPSSKTTVASEKPSIMGSI 1057

Db 914 KEVARLETLVRKAGEQOETASRELIVKEPARAGDROPEWL----- 952  
Qy 1058 SOGTPCTVLTSHNOASYTOETPKPSVGSISLGLPROOBSAKSATLPIYIKQEEFSPRNS 1117  
Db 953 -EEOGROFCTQALQAMEREAEQNGN-----ELERLRALM-----ESQG 994  
Qy 1118 QPEGLLVRAQHEGVVRGAGAIQEGSITRGPTTKISVIESIPSLRGSTOQTALPOTGI 1177  
Db 995 QOE---ERGOQE---REVARLTQE---RGAQADLALEK---AARAEI----- 1030  
Qy 1178 PTEALVKGSIEMPIEDSSPKGEBEAAASKHVITYEGSKSHLSYDNTKNAREGTRSPRT 1237  
Db 1031 --EMRLQNALNEQVEFAT-----LQEAHALHTEKEGRD----- 1063  
Qy 1238 AHEISLKRYSVEGNIKMGMSRSPSAPLEGLICRALPRGSPHSDLKERTVLGSGSIM 1297  
Db 1064 -QELAKLRGLEAQ--IKELELRQT-VKQLEQLA-----KKEKEHAGSGA 1107  
Qy 1298 QGTFRATTESFEDGLKPKQIKRESPIRAFEGAITKGPYDITTIKEMGRSIEHPRQ 1357  
Db 1108 QSEAAGRTEP-----TGPKLEALRAEVSK----- 1131  
Qy 1358 DILQESKTEPVVQSTRPIIEGSIQGTPIKFDNNSQSAIKHNV--KSLITGPSKLRS 1415  
Db 1132 --LEQCOQOQEQADS-----LERSLEAERASRAERDSALETLOGLEKAEQELGHQSAL 1185  
Qy 1416 GMPPLEIVPENIKVVERKYED-----VKAGETVRSRHTSVVSSGPSVRLRTLHEAPKAQ 1470  
Db 1186 ASAQBELAFTKQVODHISKADEKWAQVARGQEAERKNLSISS----- 1229  
Qy 1471 LSPGIYDDTSARRTPVSYQNTMSRGSPMMNRTSDVTIPPNKSTNHERKSTLTPTQRESIP 1530  
Db 1230 ---LEEEVS-----ILNR---QVLEKEGESKELKRLVMAESES-- 1262  
Qy 1531 AKSPVGVDPVVSHPSPDPHURGSTAGEVYVWSHLPTQLDPAMPFHRALDPAA-----AA 1584  
Db 1263 -----QKLEESACCRRQRPATVPQLQAA 1287  
Qy 1585 YLFOROLSPTGPGSOYQVYAMENTROTILNDYITSOQMNVNLRDPDVARGLSPREOPLGL 1644  
Db 1288 LLCGRCRAS--GRAEKQVASENLQELTSQAERAEELQELKAWQEK--FFQKEQALS- 1344  
Qy 1645 PYPATRGIDLTNMPPTILVPHPGGTSTPPMDRITYIPGTQITPPRPNYASMSPGHPT 1704  
Db 1345 -----TLQLEH---TSTQAL-----VSELLPA--K 1364  
Qy 1705 HL-----AAAASAREREREKERE-----RERIAAASDL-----YLRPGSEQPGRPGS 1750  
Db 1365 HLCQOLQAQAAAEKRRHELEQSKQAAGGLRAELLRQAQRELGLIPLURKQVAEQRTAQ 1424  
Qy 1751 HGYVRSPPSVRTOETMLQORPSVFQGTNGTSVITPLDPTAQLRIMPLPAGGPSISQGLP 1810  
Db 1425 Q--LRAEKASVAEOLSMUKAHGLIAEEN-----RGL- 1454  
Qy 1811 ASRYNTAADALAALVDAASAPQMDVSKTESKHEARLEENLRSRSAVSEQQOQLEOKT 1870  
Db 1455 GERANLGRQFLEVELDQA-----REKYVOELAAVRADAETRLAEVQREQAQSTARE 1504  
Qy 1871 LEVEKRSVOCLYTTSAPPSGKPPQPHSSVYSEAGKDGPPPKPYEEELRTRGRKTTITAA 1930  
Db 1505 LEV-----MTAKYEGAKVKVLEERQFQEB---RQKLLTAQVE 1538  
Qy 1931 NFIDVITROIADSKDARE-----RGSSDSSSLSSHRYETPSPDAIEVISPASSPAP 1984  
Db 1539 E-----LSKKLADSDQASKVQOQKLVAVQAQGESQEAQRF----- 1575  
Qy 1985 PQEKLTQYQPEVWKAQENDDPTROYEGPLHHY-RPOQESPSQOQLPPSPSQ----- 2035  
Db 1576 -QAQNELQAGLSQKEQAEEHYKLQMEKAKTHYDAKQOQELQELSLQLOKQENKEL 1634  
Qy 2036 ---AEGMGQVPRTHRLITL-ADHTCQIITODF---ARNQVSSQTPQPPPTSTFQNSPSALV 2089

Db 1635 RAEARLGHQLQAGLTKAEQTCRHLHTAQRVSRLEAQVAHADQQLRDLGKFQVATDALK 1694  
Qy 2090 STPVRTTNSRYSPESQAQSVHHQRPGRSVRPNLVDRSGSRGKSPERSHVSSEPYEP 2149  
Db 1695 SREPOAKPQLDLSLDSLDLSCGEGTPLSITS-----KLPTQDPDGTSPVCEPASP 1744  
Qy 2150 IS---PPOV-----PVVHEKQ-----DSL---LLLSQGAEPAEQORN----- 2180  
Db 1745 ISQRLPPKVESLESLYFTPIPARSQAPLESILSDSLGDFVLDGSRKTRARRTTQIINIT 1804  
Qy 2181 -----DARSPGISVLPSPFT-----KLENTSPWVKSKQEIFRKLNSGGGSDMAAAP 2231  
Db 1805 MTKKLDVEEPSAN--SSFYSTRSAPASQASLRATSSQSLAR-LGSPDYGNSALLSL-P 1860  
Qy 2232 GTEIFNLPAVTTSGSVSRG--HSFADPASNLGLEDIIRKALMGSPDDKVE-----DHGVV 2285  
Db 1861 GYR---PTTSSARRSQAGVSSGAPGORN-----SFYMGTCODEPEQLDDWNRAE 1908  
Qy 2286 MSQPMGVVPGTANT-----SVVTSGETRREEDPSPHSGVGCVCKPKLISKSNSRKS 2335  
Db 1909 LQQRNRVCPPHLKTCPYLPESRPSLSLGTITDEMKTGD-----QETLRRAS 1955  
Qy 2336 KSPI---PGQVILGTERPSSVSVHSEGDYHROTPGHAWEDRPSSTGSGTO-PPYNPLPMR 2391  
Db 1956 MQPIQIAEGTGITTRQQRKRVSLEPHQG-----PG-----TPESKKATSCFP-RPMTPR 2003  
Qy 2392 -----MLSSSTPPTPIACAPSAVNOAAHQ 2415  
Db 2004 DRHEGRKQSTTEAOKKAAPASTKQADRQ 2032

## RESULT 31

US-08-470-950-4  
; Sequence 4, Application US/08470950  
; Patent No. 5698439  
; GENERAL INFORMATION:  
; APPLICANT: TOURKATLY, GARY  
; APPLICANT: LIDGARD, GRAHAM P  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,950  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-470-950-4



|    |      |                                                               |      |
|----|------|---------------------------------------------------------------|------|
| Db | 1576 | -QAQNLNELQAOLSQEAAEHYKLMQETAKTHYDAKKQONQELQBLSLBQLOKKNEL      | 1534 |
| Qy | 2036 | ---AEGMGQVPTHRLITL-ADHICQIITQDF---ARNQVSSQTQQPPTSTFQNSPSALV   | 2089 |
| Db | 1635 | RABAEERLGHLEQAAGLKTEABQTCRHLITAQVRSLEAQVAHADQQLRDLGKFQVATDALK | 1694 |
| Qy | 2090 | STPVKRTSNRYSPESQAQSVHHQPGSRVSPENLVDRSGRSRCKSPERSHVSSPEY       | 2149 |
| Db | 1695 | SREPQAKPQLDISIDLSLDCSEEGTPLSTS-----KLPTQDPDGTGVSVPGEPA        | 1744 |
| Qy | 2150 | IS---PPQV-----PVVHEKQ-----DSL-----LLLSQRGAEPAEORN-----        | 2180 |
| Db | 1745 | ISQRLPPKVESLESIVFTPIPARSQAPLESSLDSLGDFVLDGSRKTSARRTTQIINIT    | 1804 |
| Qy | 2181 | -----DARPGSISYLPSTFT-----KLENTSPWKSKQOEIFRKLNSGGGSDMAAAQ      | 2231 |
| Db | 1805 | MTKKLDVEEPPDSAN---SSRYSTRSAPASQALRATSTQSLAR-IGSPDYNSALLSL-P   | 1860 |
| Qy | 2232 | GTBEIFNLPAVTTSGSVSSRG-HSFADPASNLGLEIIRKALMGSFDDKVE-----       | 2285 |
| Db | 1861 | GYR- ---PITRSSARRSQAGVSSGAPPGRN-----SFYMGTCODEPEQLDDWNRIAE    | 1908 |
| Qy | 2286 | MSQPMGVVPGTANT-----SVTSGETRREEDPSPHSGGVCKPKLISKNSRKS          | 2335 |
| Db | 1909 | LOORNVCVPHLTKYCPLESRPSLSLGTITDEMKTGDP-----QFTLRRAS            | 1955 |
| Qy | 2336 | KSP1---PCQGYLGTGRPPSSVSVHSEGDYHRQTPGAWEDRPSSTGSTO-FPYNPJLTM   | 2391 |
| Db | 1956 | MQPIQIAEGTGITTRQORRVSLEPHQG-----PG-----TPESKXATSCFP-RPMTP     | 2003 |
| Qy | 2392 | -----MLSSTPPTPIACAPSAVNQAAPHQ                                 | 2415 |
| Db | 2004 | DRHEGRKQSTTEAQKKAAPASTQKADREQ                                 | 2032 |

## RESULT

```

US-08-467-781-4
; Sequence 4, Application US/08467781
; Patent No. 5780596
; GENERAL INFORMATION:
; APPLICANT: TOUTKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,781
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESO, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```





Db 1505 LEV-----MTAKYEGAKVKVLEERQRFQEE---RQKLTAAQVE 1538  
Qy 1931 NFIDVITRIQIASDKDARE-----RGSQSDSSSSLSHRYETPSDAIEVISPASSAP 1984  
Db 1539 E-----LSKKLADSOAKSQVOQKLVKAVQAGGSQQAQRF----- 1575  
Qy 1985 POEKLQTVQPEVVKANQANDPTROYEGPLHY -RPOQESPSQOQLPPSSQ----- 2035  
Db 1576 -QAQLNELQAQLSQKEQAAYHKQMEKAKHYDAKKAQKQNLQELQKLSLQKQENKEL 1634  
Qy 2036 ---AEGMQQVPRHLITL-ADHICQIITQDF--ARNQVSSQTFOQPPSTFTQNSPSALV 2089  
Db 1635 RAEAEPLGHELOQAGLTKAEQTCRHLTAQVRSLEAQVAHADQQLDLGLGKQFQVATDALK 1694  
Qy 2090 STPVRTKTSNRYSPESQAQSVHQRPGSRVSPENLVDKSRGSPCKSPERSHVSSEVPEP 2149  
Db 1695 SREFOAKQDLDSIDSLDSCIEGTPLSITS-----KLPTQDGTSVPGEPASP 1744  
Qy 2150 IS---PPQV-----PVVHEKQ-----DSL-----LLLSQRGAEPAEQNR----- 2180  
Db 1745 ISQRLPPKVESLESILYFTTPARSAQPLESSILSDSLGDVFLDSGRKTRSAARRTTQIINIT 1804  
Qy 2181 -----DARSPGISYLPSSFT-----KLENTSPMKVSKQETFRKLNSSGGGSDMAAAQ 2231  
Db 1805 MTKKLDVEEPSAN--SSFYSTRSAPASQASLRATSTQSLAR-LGSPDYGNALLSL-P 1860  
Qy 2232 GTEIFNLPAVTTSGSVSSRG-HSFADPASNLGLEDIIRKALMGSFDDKVE-----DHGVV 2285  
Db 1861 GYR-----PTTSSARRSAGVSSGAPGGRN-----SFTYMGTCQDPEQLDWNRIAE 1908  
Qy 2286 MSQPMGVVPGTANT-----SVVTSGETRRREGDPSPHSGGVCKPKLISKSNRKS 2335  
Db 1909 LQQRNRVCPPLKTCYPLESRPSLSLGTITDEEMKTGDP-----QETLRRAS 1955  
Qy 2336 KSPI---PCQGYLGTSPSSVSVHSEGDYHROTPGHAWEDRPSSTGSTQ-PPYNPLTMR 2391  
Db 1956 MOPIOIAEGTGITTRQQRKRVSLPHQO-----PG-----TPESKATSCFP-RPMTPR 2003  
Qy 2392 -----MLSSSTPPTPIACAPSNAVQAAPHQ 2415  
Db 2004 DRHEGRKOSTTEAQKKAAPASTKQADRQ 2032

RESULT 34  
us-08-056-200-94  
; Sequence 94, Application US/08056200  
; Patent No. 5616500  
; GENERAL INFORMATION:  
; APPLICANT: Steinert, Peter M.  
; APPLICANT: Lee, Seung-Chul  
; APPLICANT: Kim, In-Gyu  
; APPLICANT: Chung, Soo-Il  
; APPLICANT: Park, Sang-Chul  
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
; TITLE OF INVENTION: Methods of Using Same  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/056,200  
; FILING DATE: 30-APR-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

; NAME: Fedrick, Michael F.  
; REGISTRATION NUMBER: 36,799  
; REFERENCE/DOCKET NUMBER: NIH054.001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (714) 760-0404  
; TELEFAX: (714) 760-9502  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1898 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-056-200-94

Query Match 2.2%; Score 284; DB 1; Length 1898;  
Best Local Similarity 18.5%; Pred. No. 4e-09;  
Matches 210; Conservative 170; Mismatches 376; Indels 382; Gaps 44;

Qy 33 RHQEEFAPDYRSSHLEVSQASQLQOQO---QOQLRRRPSLLSEFHPGSDRPOERTSY 89  
Db 397 RREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRRE 449  
Qy 90 EPFHPGPSVDHDSLESKRP-----RLEQVSDSHFQVSAAVLPLVHPLPEGLRASA 141  
Db 450 EERHEQKHEQERREOQLRREOQERRDWLKRTEETERHEQE-----RRKQ 493  
Qy 142 DAKDPAFGGKHEAPSPISQPCGDDONASPSKLSKEELIQSDRVDRETAKEVQOILK 201  
Db 494 QLRDQE-----EERRERWLKEEERREOQERREOQLRREOQERRE 535  
Qy 202 LKKQOQLEEEAAKPEKPPVPEQKHRSIVQIYDENRKKAEAKHIFEGLGPKV 261  
Db 536 QRLKQEEERLQRLSEOQLRR---EQEER-LEQLLKREERKRLQERR----- 582  
Qy 262 ELPLYNQPSDTKYVHENIKTNQVMRKLILFFKRRNHARKQ---REQICORYDOLMEAW 318  
Db 583 -----EQLKREOQERRDQL--KREERROQLKREO--ERLEQRLK-- 622  
Qy 319 EKKVDRIENPR---KAKES-----KTREYVEKQFPETKROQOER----- 358  
Db 623 REEVERLEQERDRERLKRREEPEERHELLKSEEEERHEQLR--REOQERERQLR 680  
Qy 359 ---FQVQGAGLGSATIAARSEHISEIIDGLSQENNEKQMLSVIPPMF-----D 409  
Db 681 EEEERLEQR-----LKREHEERREOQLAEEOQAEERIKSRIPKQWOWLESEAD 732  
Qy 410 AEQRRVFINNGLMEDP-----MKVYKDRQFMVWTDHEKEIFKDKFTQHPKNF 459  
Db 733 ARQSKV-----LLEAPQAGRAEAPQOEKRRRESELOWQEERAHROQO----- 777  
Qy 460 GLIASYLERKSVDPCLVLYYLTKKNENYKALVRNRYKRRGRNQOIARPSOEKVE--- 515  
Db 778 -----EEQRDRFTWQOAEKSE-----RGRQLRSARPLPQERERQLR 817  
Qy 516 -----EKEEDKAEKTEKKEEEKK-----DEEEKDEKSKENTKEKDIDGTA 558  
Db 818 AEERQOERQFLPEEEKEQGRGRERREKELQFLLEEEOQLORREOQLQEE--DGLQ 875  
Qy 559 ETEEREQATPRGRKTA-NSQGRKRGKITRMTWNEAAAAAASAAAAATEEPPLPPPEP 617  
Db 876 EDQERRRQERDQKWRQWLEERKR-----RHTLYAKPALQOQLRKEOQLQEEEEE 929  
Qy 618 ISTEPTVTSRWTEEMEVAKKGLVEHGRNAAIAKAVGTGKSEAOCKNFYFNKRRHLDN 677  
Db 930 LQREERKRRRQEROYREEEQLEQEE-----QLREERKRRRQ-----EREQYRKDK 981  
Qy 678 LLQO-----HKQTSRKPREERDVSCQESVASTVSAQEDIEDASNEENPE 724  
Db 982 KLAQKEQLLGEPEKRRRQREKRYREEEELQO-----EEQOLLREERKRRQ 1031  
Qy 725 DSEVEAVKPSDSDSPENATSRGNTEPAVELEPTTETAPSTSPSLAVPSTKPAEDSV--- 780



Db 1032 EWERQYRKDELQEE-----EQLLREERKRLQERQYREBELQEEELQEE 1084  
Qy 781 -ETQVNDISAEATBOMDVDOQHSAGSVCDPPATKAUSDVVEVRPE----- 830  
Db 1085 RETRRQLEQYRKDELQEE-----EQLLREERKRLQERQYREBELQEEELQEE 1140  
Qy 831 -----NHASKVEDNTKRDLDRASEKVEPRDEDLVVAQINNAQRP-----EPQSDN 877  
Db 1141 LREERKRRQLEQYREBELQEEELQEE-----QSDLDKQWQPEKEN 1191  
Qy 878 ---DSSATCSADE-----DVGGEPRORM-----F 899  
Db 1192 AVRNKVKCKRENEQFQLEDQSDQVDRQSQODLQHLLEGEQERDRERRRQWQANRHF 1251  
Qy 900 PMDSKPSLLNPTGSLVSSPLKPNPLDLPOLQHRAAVIPPMVCTPCNIPGTVPVGYAL 959  
Db 1252 PEEQ-----LEREQKEA----- 1265  
Qy 960 YORHIKAMHESALL-----EQRQROEQIDLECRSSTSPCGTSKSPNREWEVLQAPAPH 1013  
Db 1266 KRDRKSOERKQLLREERKRRQEE-----TDKRFREELQLEQEEQ 1309

## RESULT 35

US-08-800-644-94  
; Sequence 94, Application US/08800644  
; Patent No. 5958752  
; GENERAL INFORMATION:  
; APPLICANT: Steinert, Peter M.  
; APPLICANT: Lee, Seung-Chul  
; APPLICANT: Kim, In-Gyu  
; APPLICANT: Chung, Soo-Il  
; APPLICANT: Park, Sang-Chul  
; TITLE OF INVENTION: Trichochyalin and Transglutaminase-3 and  
; TITLE OF INVENTION: Methods of Using Same  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/800,644  
; FILING DATE: 14-FEB-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/056,200  
; FILING DATE: 30-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fedrick, Michael F.  
; REGISTRATION NUMBER: 36,799  
; REFERENCE/DOCKET NUMBER: NIH054.001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (714) 760-0404  
; TELEFAX: (714) 760-9502  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1898 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-800-644-94

Query Match 2.2%; Score 284; DB 2; Length 1898;  
Best Local Similarity 18.5%; Pred. No. 4e-09;

Matches 210; Conservative 170; Mismatches 376; Indels 382; Gaps 44;  
Qy 33 RHQEFAYPDYRSSHLEVSQASQLLQOOQ---QQQLRRRPSLLSEFHPGSDRPPQRRSY 89  
Db 397 RREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRRE 449  
Qy 90 EPFHPGPSVDHDSLESKRPE-----RLQVSDSHFQVSAAVLPLVHPLPEGLRASA 141  
Db 450 EERHEQKHEQERREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLR 493  
Qy 142 DAKDPAFGGKHEAPSPISQPCGDDQNASPSKUSKEELIQSDMDVRDREIAKVFQILK 201  
Db 494 QLKRDQE-----EERRERWLKLEERERREQLRREQLRREQLRREQLRREQLRRE 535  
Qy 202 LKKQOQLEEEAAKPEPEKPVSPPPVEQKHSRISQIIVYDENRKKAEAHKIFEGLPKV 261  
Db 536 QRLKQEEERLQRLRSEQLR---EQEER-LQLLKRREEKRLQERR----- 582  
Qy 262 ELPLYNQPSDTKVYHENIKTNQVMRKILILFFKRNRHARKQ---REQICORYDQIMEAW 318  
Db 583 -----EQLKREQEERDQLL--KREERERQRLKREQE--ERLEQLRK-- 622  
Qy 319 EKKVDRIENPPRR---KAKES-----KTRYIEKQFPETRKQREQER----- 358  
Db 623 REEVERLEQEEERDRERLKRPEEERHELLKSEQEERHQLR--REQOERREQLR 680  
Qy 359 --FORVGORGAGLSATIRSEHEISEIIDGLSEQENNEKQRLSVIPPMF-----D 409  
Db 681 EEEERLEQR-----LKREHEERREQLAEEOQAREIKSRIPKQWQLESBAD 732  
Qy 410 AEQRRVKFINMGLMEDP-----MKVYKDRQFMVMTDHEKEIFKQFIQHKNF 459  
Db 733 ARQSKV-----LLEAPQAGRAEAPQEQEKKRRESELOQWEEERARHQOQ----- 777  
Qy 460 GLIASYLERKSVDPCLVYLYTKNKNYKALVRRNYGKRRGNQOIIARPSQEEKVE---- 515  
Db 778 -----EEQRRDFTWQAEKSE-----RGRQLRSARPLRQERQRLR 817  
Qy 516 -----EKEEDKAETKEKEEK---DEEKEDEKESKENTKEKIDGTA 558  
Db 818 ABERQORQRLPEEEERQGRQREKEQLFLEEEQQLRRERAOQLQEE--DGLQ 875  
Qy 559 ETEEREQATPRGRKTA-NSQGRKRGRTSRMTNEAAAAAATAEPPPLPPPPPEP 617  
Db 876 EDQERRRQRRDQKRWQLEERKRR-----RHTLYAKPALQELRKEQLLQEEEEE 929  
Qy 618 ISTEPEVTSRWTEEMEVAKGLVEHGRNWAIAKMWGTKEAQCKNFYFYKRRHNDN 677  
Db 930 LQREERKRRRQEQERYREBELQEE-----QLLREERKRRRQ---ERERYRKDK 981  
Qy 678 LLQO-----HKQTSRKPREERDVQSCEVASTVSAQEDIEDASNEENPE 724  
Db 982 KLOQKEEQLLGEPEKRRRQREKRYREELQO-----EEQLLREERKRRRQ 1031  
Qy 725 DSEVEAVKPSDPSENATSRGNTPEPAVELEPTTETAPTSPSLAVPSTKPAEDSV--- 780  
Db 1032 EWERQYRKDELQEE-----EQLLREERKRLQERQYREBELQEEELQEE 1084  
Qy 781 -ETQVNDISAEATBOMDVDOQHSAGSVCDPPATKAUSDVVEVRPE----- 830  
Db 1085 RETRRQLEQYRKDELQEE-----EQLLREERKRLQERQYREBELQEEELQEE 1140  
Qy 831 -----NHASKVEDNTKRDLDRASEKVEPRDEDLVVAQINNAQRP-----EPQSDN 877  
Db 1141 LREERKRRQLEQYREBELQEEELQEE-----QSDLDKQWQPEKEN 1191  
Qy 878 ---DSSATCSADE-----DVGGEPRORM-----F 899  
Db 1192 AVRNKVKCKRENEQFQLEDQSDQVDRQSQODLQHLLEGEQERDRERRRQWQANRHF 1251  
Qy 900 PMDSKPSLLNPTGSLVSSPLKPNPLDLPOLQHRAAVIPPMVCTPCNIPGTVPVGYAL 959  
Db 1252 PEEQ-----LEREQKEA----- 1265

QY 960 YORHIKAMHESALL-----EQRQROEQIDLECRSSTSCPGTSKSPNREWEVLQAPHQ 1013  
Db 1266 KRDRKSEKQELLREERERROE-----TDRFREELQREERQ 1309

RESULT 36  
US-08-227-536-2  
: Sequence 2, Application US/08227536  
: Patent No. 5658784  
: GENERAL INFORMATION:  
: APPLICANT: Eckner, Richard  
: APPLICANT: Ewen, Mark  
: APPLICANT: Livingston, David  
: TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION  
: TITLE OF INVENTION: FACTOR P300 AND USES OF P300  
: NUMBER OF SEQUENCES: 13  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
: STREET: Ten Post Office Square  
: CITY: Boston  
: STATE: MA  
: COUNTRY: US  
: ZIP: 02109  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/227,536  
: FILING DATE: 14-APR-1994  
: CLASSIFICATION: 436  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Williams Ph.D., Kathleen A.  
: REGISTRATION NUMBER: 34,380  
: REFERENCE/DOCKET NUMBER: DFCI-308XX  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (617) 542-2290  
: TELEFAX: (617) 451-0313  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 2414 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
US-08-227-536-2

Query Match 2.2%; Score 284; DB 1; Length 2414;  
Best Local Similarity 19.0%; Pred. No. 5.6e-09;  
Matches 454; Conservative 300; Mismatches 864; Indels 774; Gaps 118;

QY 117 SHFQVSAVPLVHPLPEGLRASADAKDPAGGKHEAPSSPGCGDDQNASPSKL 176  
Db 401 SHWKNCTRHDCVCLPLK---NAGDKRKNQBPILTCAPVGLGNPSS---LGVGQDSAPNLS 454  
QY 177 SKEELIQMDRVREIAKV-----EQILKLLKKKQOOLEEAAKPEPEKPSVSP 226  
Db 455 TVSQIDPS--SIERAYALGLPYQVONMPTQPVQAKNQOQCGSQPGQGRPMNSNAS 512  
QY 227 PVEQKHRSIVO-----TIYDENRKKAEAHKIFEGGLGPKVPELPLYNQPSDT-- 272  
Db 513 PMGVNGGVGVQVTPSLSDSLMSLHNSQNPMMSENAS--VPSLGP---MPTAAQPSITGI 567  
QY 273 -KVYHENIK---TNOVVRKKLILFFKRRNHARKOREKICORYDQIME---AWKKV--D 323  
Db 568 RKQWHEITQDLRNLHVKLVQAIPTPDPAALK-----DRRMENLVAYARKVEGD 618  
QY 324 RIENPRRKAKESKREYI-----EKQFPEIRKQEQERF----- 359  
Db 619 MYESANNR-----AEYHLLAEKIYKIQELEKEKRTRLRQKONMLPNAAGWVPSMNP 671

QY 360 -QRYGQRGAGLSATIRSEHEI-----SEIIDGLSBOEN-----NEKOMROLSVI---P 404  
Db 672 GPNMGQPOPGMTNGPLDPDSMIRGSPVQNMWPRITPQSGLNQFGQMSMAOPPVPRTPT 731  
QY 405 PMMFDAGORRVKFTINMGLMEDPMKVYKDRQFMVWTDHEKEIFKDFIQHPKKNFLIAS 464  
Db 732 PLOHHGQ-----LAQPCALNPPMG-YGPR-----MQQPSNOG---Q 763  
QY 465 YLERKSVDPDCVLYYYLLTKKNENYKALVRNNGKRRGNQ---QIARPSOEKVEKEEDK 521  
Db 764 FLPTQTFPS-----OGMNVNIPLAPSSGQAPVSOAQM-- 796  
QY 522 AEKTEKKEEKDEEKEDEKDKIDCTAEETEEREQATPRGKTANSQGR 581  
Db 797 -----SSSSCPVNSPIPPG-----SQGSH 816  
QY 582 KGRITRSMTNEAAAAAATAEPPPPPLPPP-----PPISTEPVETSRMT 629  
Db 817 IH--CPQLPQALHONSPSPVPSRTPTPHHTPPSIGAQPPATTIPAPVPTPP----- 867  
QY 630 EEMEVAKKGLVEHGRNWAIAKMGVTKSEAOCKNFYENYKRRHNLNLLQOHKQKTSRK 689  
Db 868 -----AMPPGQSQA-----LHPPPRQPTTP 888  
QY 690 PREERDVSQCESVASTVSAQDEDEIEASNEEENPDESEAVK-PSEDS-----PENATSR 744  
Db 889 PTTQLPQVQVPSLPAAPSADQ-----PQQPRSQQSTAASVPTNAPLLPQPPATPL 940  
QY 745 GNTPEPAVELEPTTETAPSTSP-----SLAVSTKPAEDESVEVQVNDISAEATQMDVQ 800  
Db 941 --SQPAVSIEGQVSNPPTSTSTEVNSQAIKQPSQEVKMEA-----KMEVDQ 986  
QY 801 QESAEGSVCDPPPATKADSDVDEVVRPENHASKVGDNTKERDLDRASEKVEPREDL 860  
Db 987 PE-----PADTQPEDIS-ESKVED---CKMESTETEERSTELKTIKEEEOQS 1031  
QY 861 VVAQOINARPEPOSNDNS-----SATCSA--DEVDGEPERQRMFMDSKPSL 907  
Db 1032 TSATQ---SSPAPQSKKIKFKPEELRQALMPTLEALYRQDPESLPFRQ---PVD--PQL 1083  
QY 908 LN-PTGSILYSSPLK---PNPLDLQLOHRAAVIPMWVCTPCNIPGTGVPVGYALYQR 962  
Db 1084 LGIPDYFDIVKSPMDLSTIKRKLDTQYQEPWQYVD-----DIWLMFNANWLYNR 1133  
QY 963 HIKAMHE--SALLEEQRQEQIDLECRSSTSCPGTSKSPNREWEVLQAPHLITNLPE 1020  
Db 1134 KTSRVYKCYKSLSEVFEQ---EIDPVMQSLGYCCGRK-----LEFSQTLCC---Y 1178  
QY 1021 GVRLPTRTRPPPLIPSSKTTVASEKPSFIMGSSISQGTGTYLTSHNOASTQETPK 1080  
Db 1179 GKQLCT-----IPRD-----ATYYSQNYRHYHCEKCFN 1206  
QY 1081 PSVG-SISLG-LPQOESAKSATLPYIKOEFEFSRSONSQPEGLLVRA----- 1126  
Db 1207 EIOGESVSLGDDPSQPTT-----INKEQFSKRKNNDLDPPELFVECTEGRKMHQICV 1259  
QY 1127 -QHEGVVGTAGATAQEGSITRGTP---SKLSVESIPSLR-GSITQG-----TPALPQ 1174  
Db 1260 LHHE--IIPWAGVCDCLCKKSAATRENKFSAKRLPSTRGLTLENRVNDFLRQNHPE 1317  
QY 1175 TGITEALVKGSISSRMPIE-----DSSPEKGREBAASKGHVYIEKSGHILSYONIK 1226  
Db 1318 SGEVTVRVVHASDKTVEVKPGMKARFVDSGEMAESFPYRTKALFAFEIDGVDLCFFGMH 1377  
QY 1227 NAREGTRSP-----RTA--HEISLKSYESVEGNIKOGMSMR 1261  
Db 1378 VQEGSDCPPPNQRRVYISYLDVSHVFFRPKCLRTAVVHEILI--GY--LEVYKKGITGT 1433  
QY 1262 ESPVSAPLEG--LTCRALPRGSPHSLDKERTVLSGSIQMGTPPRATTE-----SFDGLKYP 1315  
Db 1434 HIWACPPSEGGDDYIFHCHP---PDQKIPKPKLQEWYKMKLDKAVSERIVHDYDKIFK-- 1488  
QY 1316 KQIKRESPPIRAFEGAITKKG--PY-DGIIFFIKEMGRSIIHEIPRODILTOESRKTPEVQ 1372





COUNTRY: U.S.A.  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/853,913  
FILING DATE: 19920319  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/660,412  
FILING DATE: 22-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/603,803  
FILING DATE: 25-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/201,762  
FILING DATE: 02-JUN-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/180,548  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-4641AAAA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1618 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-853-913-4

Query Match 2.28; Score 276.5; DB 1; Length 1618;  
Best Local Similarity 19.58; Pred. No. 9.3e-09;  
Matches 316; Conservative 215; Mismatches 521; Indels 565; Gaps 83;

Qy 10 QGAFSTEGSRYPHPSVQVTFNTRHQOEFAVDPYRSSHLEYSQASQLQQOQQQLRRRP 69  
Db 186 RGAAGYQYERVA--HMETSLDQTRERLARAVOGAREVLELQ--LQAEKGGLLERRA 239  
Qy 70 SLLSEFHGSDRPRQERTSYEPFHGSPVDHD----- 102  
Db 240 ALEQRL-----EGRWQERLATEKFKQLAVEALEQEKQGLQSQIAQVLEGRQQLAHLKMSLSL 296  
Qy 103 -----SLESKRPRLE-QVSDSHFORVSAVLPLVHP--LP 134  
Db 297 EVAYRTILLEANSRLQTPGGGKTSLSFQPKLEQPRTPPEGRRLGSLPLVLSPTSLP 356  
Qy 135 EGLRASADAKKDPAGFKGHE-----APSSPIGQPCGDDQNASPSKLSKEELIQSMD 186  
Db 357 SPLPATLETVP-PAFLNKQEFQARTPTLASTPIPTP-----QAPSPA----- 399  
Qy 187 RVDREIAKVEQIQLK-----KKQOOLEEA--AKP-----PEPEKP-----VS 224  
Db 400 -VDAEIRAQADAPLSLLQQTQGRKQAPLEPLRAEAVIPASVLPPEPGGQREASTQGS 458  
Qy 225 P-----PPVEQKH-----RSI 235  
Db 459 PEDHASLAPPLSPDHSLEAKDGSGSRVFSICRGEQGIWGLVEKETAIEGKVYSSL 518  
Qy 236 VOIYDE---NRKAEAEH-----KIFELGPKVELPLYNQPSDTKVYHENIKTNQVMRK 287  
Db 519 QOEIWEEDLNARKEQDSQVPLEKETLSLGEIEQ-----ESLKT----- 558  
Qy 288 KLILFFRRNRHARKQREKQICQY-----DQLEWAEKKVDRI-----ENNPRKAKES 336  
Db 559 -----LENQSHETLERNEQCPRLSEEDLETLSLEKENKRAIRKCGGSETSRKRCRQL 613

Qy 337 K-----TREYVEKOPPEIKKORE-QOERFORVQORGAGLSATIAARSEHSEISIDGL 387  
Db 614 KPTGKEDTQTLOSLOKENQELMKSLGNLETFLFPGTENQELVSSLOENLESITLAL----- 669  
Qy 388 SEQENNEKOMRQLSVIPPMFMDAEQRRYKFINMGLMEDPMKVKYKDRQFMNVTWTHKEI 447  
Db 670 -EKENQE-----PLRSPVGVDEALR-----PLTKENQEPLRSLR-----DENKEA 709  
Qy 448 FKDFIQHPKNEGLIASYLERKSVDCVLYYYLTTKN-ENYKALVRYNRYKRRGRNQOIA 506  
Db 710 FRS-----LEKENEQPLKTL-----EEDQOSIV 732  
Qy 507 RPSOEE-----KVEEEDKAETKEKEEKKDEEKEDEKEDSKENTKEDKIDGTAAE 560  
Db 733 RPLETENHKSLSRLEEQQOETLRTLEKETQQR-----RSLGEQDQMTLRPPKVD--LEP 786  
Qy 561 TEEREQATPRGRKTANSQGRKGRITRSMTNEAAAA-----SAAAAATEPPPPPL 611  
Db 787 LKSLDQEIARPLENQE-----FLKSLKEESVAVKSLTEILESLKSAGQEN----- 835  
Qy 612 PPPPEPISTEPVETSRWTEEM-----EVAKKG-----LVEHGRNWAATA 651  
Db 836 ---LETLSKETQAPLWTPPEINKSGNNESSRKGNSTRITGVCGSEPRDIQTGPGESGII 892  
Qy 652 KMGVTKSEAOCKNFYFNKR-----RHLN---DNLLOQHKOCTSRKPREE-----RD 695  
Db 893 EISGSMERGE-----FEISRGVDKESQNRNLEEEENLKGQEQESLRSLEEGQELPQSD 947  
Qy 696 VSQCESVASTVSAQEDDI-----EA-----SNEEENPEDEVEEA 730  
Db 948 VORWED-----TVEKDQELAQESPPGMAGVENKDEAELNLRQDQGTCKEEVVEQELNA 1002  
Qy 731 VK-----PSDESPEN-----ATSRGNTEPAVELETTETATPTSPSLAVPSTKP 774  
Db 1003 TEVWFPGEHPENPEKPEQRGLVEGASVKGGAEL--QDPGQSQQVGTGGLQAQGLP 1060  
Qy 775 AEDSESVTOVNDISIA---ETAQOMDVQOEHSABE-----GSVCDPPPPATKA 819  
Db 1061 ---EAIPELVEDDVAGDQDQASPEVNLGSEFAMGSAAGAPGLGQGVGGDGPGLHIRE 1117  
Qy 830 DSDVDEVVRPNHASKVBDNTKERDLDRASEKVEPRDEDLVAAQOINAQR--EPQSDN 877  
Db 1118 EVMPPELSEESLEAKRVQGLEGRPKDLEAGLGTFESE--LPGKSROPWEPPREGRES 1175  
Qy 878 DSSATCSADEVDGEPERQRMPPMDSKPSLNPCTGSLVSSPLKPNPLDLQLQRAAVI 937  
Db 1176 BAEAPRGABE-----APPAET-----LGTGSD-----AFSPWPLGSEAEEDVP 1215  
Qy 938 PMVSVCTPCNIPi-----GTPVSGVALYORHIKAMHES-----ALL 973  
Db 1216 PVLVSPSTYPTIILEDAPGLQPAEGSQEASGWQGRAEAKGVSEQEELSGEIPGLO 1275  
Qy 974 EQRQRQEQIDLECRSSTSPCGTS-----KSP-NREWEVLQAPHQLITNLPEGVRLPTT 1027  
Db 1276 EGEESRESEDEELGTLPDSTPLGFLYLRSPTSRPTWPLESRGHPKLTKEGK-- 1330  
Qy 1028 RPTRPPLPLIPSSKTTVASEKPSFMG-----GSISQ-----GTPGYTLTSHNQASYT 1075  
Db 1331 ---DPAVLASBGLPEPSEKEEGEEGECROSDLSEEFDLGTEAPFL-----PGVP 1380  
Qy 1076 QETPKPSVGSISLGL-----PROQESAKSATLPYIKOEFSRSONSQPEGLLVRAQHE 1129  
Db 1381 GEVAEP-LGQVPPQLLLDPAANDROGESD-----GFADSEESGEEGEEQEEG--REP 1429  
Qy 1130 GVVR-----GTAGAIOE-GSITRGPTTSKISVE-SIP-----SLRGSTQGTPTALPQTGIPTE 1180  
Db 1430 GAGRWGPGSSVGSILQALSSQRFLESDSVSVVPWDDSLRGAVA---GAPKTALETE 1485  
Qy 1181 ---ALVKGSLSRM-PI-----EDSP-----EKGEEEAASKGHVI-----YEKKSCH 1218  
Db 1486 SDSAEPGSGSEESDPVSLEREDKVPGLPTEIPSGMEDAGPGADIIGVNGQGNLEKGSQH 1545



|    |      |                                                                |      |
|----|------|----------------------------------------------------------------|------|
| Db | 1909 | TKLEPSEDGCKPELLEEEMASSTELIAVEGTIELQDFONKTDGOVSGEALMKFFITRPE    | 1968 |
| Qy | 2199 | ENTSPVMVKSKOBIFRKLNSGGSDMAAAQPGE-----IFNLPAVTITSGSVS           | 2248 |
| Db | 1969 | AGT--VIITADEIELE-CATOWPHSTASATYGVGAVPWLSQPTSERPTLSSSEPIN       | 2024 |
| Qy | 2249 | S-----RGHSFADPASNLGLIEDIRKALMGSEFD----DKVEDHGVVMQPMGVVPSTA     | 2297 |
| Db | 2025 | PETQAALIRQGDSITAAS---EQOVAARILDSNDQATVNPFVENTEVATPPFSLLTSN     | 2080 |
| Qy | 2298 | NTSWVTSGTREEGD---PSPH-----SGGVCKPKLISKSNRKSKSPITGGGYL          | 2345 |
| Db | 2081 | ETDFLGINEEVESGETAIYLPQGDRCKNPCPLNGSTCYP-----TETSUYCTCYP--GYS   | 2133 |
| Qy | 2346 | GTERPSSVSIVHSEGDIYHRQTPGWAWEDRPSSTGSTQFP-YNPILTMRLMSLTPTPTIACA | 2404 |
| Db | 2134 | GDQCELDFFECHS-----NPCRNGATCVDGFN--TFRCL-----CL                 | 2167 |
| Qy | 2405 | PSAV 2408                                                      |      |
| Db | 2168 | PSYV 2171                                                      |      |

```

RESULT 40
US-08-459-568-2
; Sequence 2, Application US/08459568
; Patent No. 5811304
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,568
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1706 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-459-568-2

Query Match 2.2%; Score 274.5; DB 2; Length 1706;
Best Local Similarity 18.0%; Pred. No. 1.3e-08;
Matches 365; Conservative 224; Mismatches 704; Indels 739; Gaps 81;
QY 502 NQIARPSQBEKVEEKDAAKTEKKBEEDKEDSKENTKEKDIDGTAEET 561

```

[illegible]





Db 554 SNISENLNYYIDGKIQNSST-----SNCVDIEMESNAHLYGIDCLLTPVVEITQ 605  
Qy 1070 N-----QASYTOETPKPSVGSISLGLPRQESAKSATLPYIKOEFEFSRSONSQEGLIVR 1125  
Db 606 NIKSTQSVTDLLKQSPSSNCE- SKKRTASPVLPKIKTETESDST----- 653  
Qy 1126 AQEGVVRGTAGATQEGSITRGTPTSKISVESIPSLRGSTQGTTPALPQTGPIPEALVKG 1185  
Db 654 -----APSCSL-----LPL----- 663  
Qy 1186 SISRMPIEDSSPEKREASKHVIYEGKSHILSYDNIKNAREGTRPTAHEISLKR 1245  
Db 664 SISTAEVVSFKKAG-----VYLSKLUQL-----QDKLTLPA 699  
Qy 1246 SYESVEGNIKQGMRESVPSAPLEGITCRALPRGSPHSLDKERTVLGSGIMQGTTPRATT 1305  
Db 700 GFSAAE-----IPKLGPCASAP-----ASMLP-----VTS 725  
Qy 1306 ESFEDGLKYPKQIKRESPPIRAFGATKKGYPDGITTIKEMGRSIIHEIPQDILTOESR 1365  
Db 726 SRFKRTSSPSPQHPALRDF-----GKPDG-----KAA 757  
Qy 1366 KTEPVOSTRPIIBEGSISOGPPKIFEDNNNGOSAIKHNKVSILITGPKSLRGMPPLEIYPE 1425  
Db 758 WTDVILTSKKP-----KLEGRSDSPA-----LSGRDERETGSPPC----- 794  
Qy 1426 NIKVVERKGYEDVKAGE--TVRSRHTSVVSGSPVLRLTHEAPKAQLSPGIYDDTSARR 1483  
Db 795 -----FDEYKISKWEAASSTFSSVCNQPP-----LDLSSGV----- 825  
Qy 1484 TPVSYQNTMRCGSPMMARTSDVTTPPNKSTNHERKSLTPTQRESIPAKSPVPG---VDP 1540  
Db 826 -----KQKSEGTGKTPVPWESVLDL 845  
Qy 1541 VVSHSPDPHHRGTAGEVYWSHLPTQDPAMPFHRALDPAAAYLQFQRLSPTPGYPQ 1600  
Db 846 SVHKKPCD-----SEGKFEKNHL-----AQPA-----KKKKPTCM--- 878  
Qy 1601 YQLYAMENTROTILNDXITSQOMQVNLRPDVARGLSPREQGLPYPATRIIIDLNNPP 1660  
Db 879 -----LQKVLLNEX-NGVSLTETITTEVTRSPSPCKSPDTQDPDL-CPDSSCSVPT 928  
Qy 1661 TILVPHCGGTSTPMWRITYIPGQITFP--RPNYASMSPGHPTHLAAASAERER 1718  
Db 929 AESPPVVGSPSPLOQASLSSG---QLPPLLTPTERPSPPCPVLTVATP----- 977  
Qy 1719 ERERERERITAAASDLYLRPGSEQRCGSGHYVRSVSPSVRTOETMLQORPSVQGT 1778  
Db 978 -----PPPLLTVPULSH-----PS-----SDASPOCCSPFSNT 1006  
Qy 1779 NGTSVITPLDPTAQLRIMPLP-----AGGP-----SISQGLPASRYNTAADA-- 1820  
Db 1007 TAQSPLPLTSTVSPSPSPIPVPEPLMSAASGPPTLSSSSSSSSSPSSCSSTSPSP 1066  
Qy 1821 -LAALVDAASAPQMDVS-----KTESKHEAARLEENLRSAAYVEEQOOLEOKTLEV 1873  
Db 1067 PLSAVSSVSSGDLLEASLPAVTEKQEESEGLKPKX-----EAPPAGGQSVVQETP-- 1119  
Qy 1874 EKRSVQCLYTSSAPSPGKQPHSSVVYSEAKDKGPPPKSYEEELTRGKTTITANFI 1933  
Db 1120 -SKNFICNVCEPSPLSKDLTKHLSVHAEEW-----PFKCFVOL-FKVKTDLSEHRFL 1172  
Qy 1934 -----DVII-----TRQIASDKDAREGSSSSSSLSHRYETPSDAIEV 1975  
Db 1173 LHVGNIFFVCSVCKKEAFLCNLQHQDLHPDEVCVTHHEFESGTLRPNQTFDSKA--- 1229  
Qy 1976 ISPASSPAPPEKLOTQYQEVVVKANQAE-----ND 2005  
Db 1230 -NVEHMPSLPEPLETSREEELNDSSELYTTIKIMASGIKTKDPVRLGLNQHPSPFKP 1288  
Qy 2006 PTRQYEGPLHHYRQQQSPS-----PQ---QQLPSPSQAEQMGQVPTHRLITLAD 2053  
Db 1289 PPQY-----HHRNPMGIGVATNFTTNIPQTFTAIRCTKCGKGVDMNPELHK----- 1338

Qy 2054 HICQIITQDFARNOVSSOTPOQPPTSTFQNSPVSALVSTPVRTKTSNRYSPESQAQSVHHQ 2113  
Db 1339 HILACASASDKRITPKKNP-VPLKQTVQPKNGVV---LONSGKNAPRMCQPKRLSFN 1394  
Qy 2114 RPSGRVSPENL-----VDSKRSRSP-----GKSPERSHVS----- 2143  
Db 1395 VELGKMSPNKILKLSALKKKNLQVOKAILQKNRAAKQADLRDTSEASHICPYCDREFTY 1454  
Qy 2144 -----SEPYEPISTPOVPVYVHEKQDSILLLSORGAEPAEQORNDARSPGISYLP 2193  
Db 1455 IGSINKUHAFCPPKPPSPSRKVSHS-----SKGGHASSSSDRNS----- 1497  
Qy 2194 FFTKLENTSPMVKSKQEIFRKLNSGGGSDMAAAQPGTEIFNLPAVTTSGSVSSRGH- 2252  
Db 1498 -----SCHPRRTADTEI--KMOSTQAPLGKTRARSTGPAQASLP-----SSSPSRQNV 1545  
Qy 2253 SPADPASNLGLEDILIRKALMGSPDDKVEDHGVMSQPMGVVPGTANTSVTSGETRRREG 2312  
Db 1546 KFAASVKS-----KKASSSL-----RNSSPIRMAKITHVEG 1577  
Qy 2313 DPSPHSGVGCKPKLISKNS-----RKSKSPIPGOGYLGTER 2349  
Db 1578 K-----KPAKAVKSHSAQLSKSRGLHVRVOKSKAVIQSKTALASKR 1620

RESULT 42  
US-08-516-859A-2  
; Sequence 2, Application US/08516859A  
; Patent No. 6069231  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/516,859A  
; FILING DATE: 18-AUG-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,411  
; FILING DATE: 06-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/292,683  
; FILING DATE: 18-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 1776  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1706 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-516-859A-2



APPLICANT: McKay, Ronald D.G.  
 APPLICANT: Lendahl, Urban  
 TITLE OF INVENTION: Nestin Expression As An Indicator of  
 TITLE OF INVENTION: Neuroepithelial Tumors  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 STREET: Two Militia Drive  
 CITY: Lexington  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02173  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/853,913  
 FILING DATE: 19920319  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/660,412  
 FILING DATE: 22-FEB-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/603,803  
 FILING DATE: 25-OCT-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/201,762  
 FILING DATE: 02-JUN-1988  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/180,548  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Granahan, Patricia  
 REGISTRATION NUMBER: 32,227  
 REFERENCE/DOCKET NUMBER: MIT-4641AAAA  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-861-6240  
 TELEFAX: 617-861-9540  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1805 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-853-913-2

Query Match 2.1%; Score 270.5; DB 1; Length 1805;  
 Best Local Similarity 19.7%; Pred. No. 2.6e-08;  
 Matches 308; Conservative 215; Mismatches 565; Indels 479; Gaps 73;  
 QY 21 PPHSVQVTFNTRHQEFVDPDYRSSHLYSVQASQLLQQQQQQQLRRR---PS---LLSE 74  
 DB 392 PTISEAPCPNNA-----EVRQAQVPLS---LLQTAPEPLWLKATVPSSAILPE 438  
 QY 75 F-HPGS-----DRQERTSPFPPHGPSPVDHDSLEKRRPRLEQVSDSHFORVSAAV 126  
 DB 439 LEEPGGQGHFPDDLTLNLPNPHPTLEAKDGESESR-----VSSIFDEGQI 491  
 QY 127 LPLVHPLLEGRASADAKDPAGFGKHEAPSPSPISGPGDQDQNASPKLSKEELIQSMD 186  
 DB 492 WELY-----EKEADIEVKVENSNA-----QKTESGLDTEETQDSQG 528  
 QY 187 RVDREIAKV--EQOILKLLKKQQLLEAAKP-----PEKPVSPPPVE 229  
 DB 529 PLOKETLALGEPLMSLK---IQNYETACKENCNSSTEGHLGTLGPEKEKQIPLKSLE 585  
 QY 230 QKRSIVQIIVDENRKAEEAHKIFEGLPKVELPLYNQPSDKVYHENTKTNQVMKKL 289  
 DB 586 EKN-----VSEKTELVGVPVLSLLGK-----EDRT-----EDQLMSPKG 623  
 QY 290 ILFFKRRNHARKOREQICQRYDQLEAWKVKVDRIENPRKAKESKTREYY----EKQ 345

DB 624 TL--KRFSSSLGKSEQEVVRPSKEGNLESWTAFKEESQPLGFGPAEDQMLERLVEKEDQS 681  
 QY 346 FPEIRKOREQO--ERFORVGORGAGLSATIARSHSEIIEIDGLSEQENNEKQMRQLSVI 403  
 DB 682 FPRSEEDQACRPLQKNOEPLGY-----EEAGQIILRIEKESE-----SLR 728  
 QY 404 PPMFADAEQRRVKFINNGLMEDPMKVYKDRQFNMVWTDHEKEIFKDKFIQHKNFGLIA 463  
 DB 729 SPEEDQEAQR---SLQKNEQPLG-YEEAEDQMLERLIEKE--SOESLAKSPENQRIG 781  
 QY 464 SYLERKSVPCDVLVYLYTKKNENYKALVRNRYGRRGRNQOIARPSQOEKVEEKEAE 523  
 DB 782 KPLERENQKS---LRYLEENQETEVPLESRN-----QRPLRSEVEEERQIVK 827  
 QY 524 KTEKKEEKKDEEKD-----EKDSKENTKCKDKIDGTAETEEREQATPRGRKTA 575  
 DB 828 PLEKVSQSLGSLAEENVQPLRYLEEDDCINKSLLEDTHKSLGSLDNRN----- 877  
 QY 576 NSGRRKGRITRSMTNEAAAAAATAEPPLPPPPPEP-----ISTEPVETSRW 628  
 DB 878 -----GDSIIIPQESQVSLRPEEDQRIVNHLEKESQEFERS 917  
 QY 629 TEEMEYAKKGLVEHGRNMAIA-----KMGYKSEAQCXNFYFNKRRINLNLQOH 682  
 DB 918 SEEEQYMERSL--EGENHESLSVEREDQMVESQLE-----KESQDSKSLDE 965  
 QY 683 KOKTSRKPRERDVYQCESVASTVSAQEDEDIE-----ASNEE---ENPEDSEVEA 730  
 DB 966 SQETFF-GPLEKENAESLRLAG--QDQEEQKLEQETOQLRAVNEQNAVSPKVDPEL 1022  
 QY 731 VKPESDPENATSRG--NTEPAVEL-EPTTETAPSTPSLAVPSTKPAED---ESVETQ 783  
 DB 1023 PKPLGNDQEIARSLGKNOESLSLKEGIEYKSLKLEIIEPLETAEDLERKKSIDTQ 1082  
 QY 784 V---NDSISAFTAEQMD-----VDQ-----QEHSAEGSVCDPPPATKADSV 823  
 DB 1083 EPLWSTEVARETVPEPDEPPGSLGSDVENRETITLSLEKESQELSSLGKWNVETRVESQ 1142  
 QY 824 VEVVVPENHASKVEGDNTKERDLDRASEKVEPRDIEDLVAAQIINARP-----EPQS 875  
 DB 1143 QCLQVEGLQEEHQESLREVYKQLPSSGNOQRWEDVVEGKAVGQEAFLATGTGVEDKA 1202  
 QY 876 -----DNDSSATCSADEDVDG-----EPERORMPMDSKPSLLNPTGILVSS 918  
 DB 1203 ELHLRGOGGEEEAAGELLODIVGEANSLGSSSEFKEQV--PAEALDNLGGALEVPVQ 1261  
 QY 919 -----PLKNPLDLPLQOH---RAAVIPPMVS 942  
 DB 1262 SMPVTERDEDRAGAQEQDSIEVTIGLEAARTGLEQEVVGLDPPRHFAREEAIPPSLG 1321  
 QY 943 CTPCNIPDTPVSGYALYQRIHAKMHESALLEQORQEQIDLECRSTSPCGTSKSPNR 1002  
 DB 1322 EESVAKIAQGLEPGKEPKEAGAL-DSGILELPKTSSEA--LEQ-----GHESESM 1372  
 QY 1003 E-WEVLOPAPHQLITNLNPEGVRLPTTRP-----TRPPPLI-PSSKTTTVA 1047  
 DB 1373 EGWEEEAAS---LETSDHEGSDAPQPRPTEDEGAQAALATPGKLLLEPCSPILTD 1429  
 QY 1048 KPSFIMGISQGPPTGYLTSHN---QASYTQET-----PKPSVSGSIS 1087  
 DB 1430 -----AHQLQPAQEGIQEAGWQPEAGSEALERVEPEFGLGEP 1469  
 QY 1088 LGLPROQESAKSATLPYIKQEEFSPRQNSQPEGLLYRA-----QHEGVVVRGTAG 1137  
 DB 1470 EGLQDWEEGREES-----EADDLGETLPDSTPLGLYLRSPASPKWDLAGQORLSPQGDAG 1524  
 QY 1138 -----AIQESITRGTP-----SKISVESIPSLRGSIQTGTALPOTGPTALV 1183  
 DB 1525 KEDWGPVAPPAAGLGGPPEEEEGCHGSDLSSEFEDL-----GTASLLPGVPKE--V 1576  
 QY 1184 KGSISRMP-----IEDSSPEKGREE-----I : : : : :

Db 1577 ADHVGPVPPVLPQACWQGGSDGFADEESERGEEDADEAGSQAQWGWGASGGG 1636  
Qy 1204 -----AASKGHVYE--GKSHILSYDNKNA-----REGTRSPRTAHSILKRS 1246  
Db 1637 CKVODIAQRGDVPQESVGVSG--LWDDGLRGAGAAANVPALEMVSQDSAPSGSEE--SESAS 1693  
Qy 1247 YESVEGNIKQGMRESPPASPLEGLICRALPRG--SPHSDLKERTVLSGSIMOGTPTAT 1304  
Db 1694 LEGEEGVTDHLDAPQE-VTSMVPG-VGDAFDIGQSPNLDSEQ---VNGKMENGLQAE 1748  
Qy 1305 TESFEDG 1311  
Db 1749 GQVVLDG 1755

RESULT 44  
US-09-041-886-23  
; Sequence 23, Application US/09041886  
; Patent No. 6235872  
; GENERAL INFORMATION:  
; APPLICANT: Bredesen, Dale E.  
; APPLICANT: Rabizadeh, Sharro  
; TITLE OF INVENTION: Protopoctic Peptides, Dependence  
; TITLE OF INVENTION: Polypeptides and Methods of Use  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/041.886  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 2626  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1185 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-041-886-23

Query Match 2.08; Score 258; DB 4; Length 1185;  
Best Local Similarity 19.68; Pred. No. 8.4e-08;  
Matches 297; Conservative 147; Mismatches 597; Indels 472; Gaps 71;  
Qy 538 KDEKSKENTKEKDIDGTAEETEEERQATPRGRKTANSQGR--RKGRIT--RSMTNEAA 594  
Db 4 RQKDSMSMRGRKKEAPGPREELRSGRASPGGVSTSSDGGKAEKSRQAKKARVEAS 63  
Qy 595 AASAAAAATEPPPPPPPEPTISTEPTVTSRTEEMEVAKKGLVEHGRNWAIAKMV 654  
Db 64 TPVKNQGRSE-----EISESESEETNAPKK----- 89  
Qy 655 GTKSEACKNFYFNYKRRHNDLNLQHKOKTSRKPREERDVSO-----CESVASTVSAQ 709  
Db 90 -TKTEQELP-----RPQSPSDDSL--DGRSLNDGSSDPDRDIDODNRKSTSPSIYSVSVE 142

Qy 710 EDEIEASNEENPEDSEVAVKPSEDSPENATSR---GNTERPAVELEPTTETATPSTPS 766  
Db 143 NDSSSSGLSGQPARPYHPPLFPSPQPDSTPRQEAEPSPHPTGTYHAPMEPT 202  
Qy 767 ----LAVPSTKPADESVETQVNDISIETAEDMDVQOEHSAEBSVCDPPPAKADSV 822  
Db 203 SRMFOAPPGAPPH-----POLYPGGTGGVLGGPMGPKGGG 240  
Qy 823 DVEVVRVENHASKVEGDNTERDLDRASEKVEPRDEDLVYAAQINAQORPEPQSDNDSSAT 882  
Db 241 -----ASSVGGPNG-----GKHPPPTTPISSVSS 265  
Qy 883 CSADEDVDGEPERQRMFPMDSKSLNPTGSLVSSP-----LKNPLDLPLQ--HR 933  
Db 266 -----GASGAP-----PTKPTTTPVGGNLPASAPPANFPHVTNLPPLPALPLNN 312  
Qy 934 AAVIPMVSTPCNIPITGTPVSGYALYORHIKAMHESALLEEQRQREQIDLECRSSTSP 993  
Db 313 ASAPPGLGAQP--LPGLHP--SPYAMGQ-----GMGGLP 343  
Qy 994 CGTSKSPNREWEVLOPAPHQLITNLPEGVRLPTTRTPRPPL-IPSSKTTVASEKPSFI 1052  
Db 344 PGPEKGP-----TLAPSPHS-----LPPASSAPAPPMRFPYSSSSSSAAAASSS 388  
Qy 1053 MGSISQGTGCTYLTSHNQASYQTETPKPSVGSISLGPLRQOESAKSATLPIYIKQEEFSP 1112  
Db 389 SSSSSSSNASP--FPASQALPSYHSPFPPT-----SLSVNQPKYTOPSLP--SQAVWSQ 440  
Qy 1113 RSQNSQPEGLLVRAQ--HEGVVRCTAGAIQEGSITRGTPTS-----KISVESIPSLRGS 1164  
Db 441 GPPPPPPYGRLLANSNAHPGPPPTGA---QSTAHPPVSTHHHHQOQOQOQOQOQ 496  
Qy 1165 ITQGTALPQTGIPTEALVKGISRMPIEDSSPEKREEAASKGHVIEGKSGH--LSYD 1223  
Db 497 QHNGSGPPPPGAPPHLEGSSSHAHPYAMSPSLSURP-----YPPGPAHLPPPHS 549  
Qy 1224 NIKNAREGTRSPRTAHEISLKRYSVEGNIKQGMRESVPS--APLEGICALPRGS 1281  
Db 550 QVSYSQAGNPPV-----SSSSSSSSTSGSYPCSHSPSQG-----PQGA 592  
Qy 1282 PHSDLKERTVLSGSIMOGTTPRAITTESFEDGLKY-----PKQIKRESPPIRAFEGAIKG- 1335  
Db 593 PYPPPPVPTVTTSATLSTVIATVASSPAGYKTASPPGPPYGRAPSPGAYKATATPGY 652  
Qy 1336 KPYGITTIKEMGRSIEHPRQDILITQESRKTPEVWOSTRII--EGSISQGTPIKFDNNS 1394  
Db 553 KP--GSPSPFTG-----TPGYRGTSPPAGPGTFKPGSP----- 685  
Qy 1395 GQSAIKHNKSLITGPSKL-----SRGMPPLEIVPENIKWVERGKYEDVKAGETVRSRHTS 1450  
Db 686 -----TVGPGPLPPAGPSGLPLPPP-----A 708  
Qy 1451 VVSSGSPVLSTL-----HEAPKAQLSPG-----YDDTSARTPVSYONTWSRG 1495  
Db 709 APASGPPLSATQIKQEPAAEYETPESVPARSPPPKVVVDVPSHASQSAFNKHLDRG 768  
Qy 1496 SPMMNRTSDVTIPP-----NKSTNHERK 1518  
Db 769 FNSCAR--SDLYFVPLEGSKLAKRADLVEKVRREAOAREEKEKEREKEREKERE 827  
Qy 1519 STLTPTQRESIPAKSPV--PGVDVPSVSHSPDPHRRGSTAGEVYWSHLPTOLDPAMPFHR 1576  
Db 828 RELERSVKLAQEGRAPVECPSLGVPVHRPPEP--GSAVATV-----PPYLGPDTPALR 879  
Qy 1577 ALDPAAYLFFQRLSP-----TPG-----TPSQYL--YAMENTROILN 1615  
Db 880 TLSEYARHV-----MSPGNRNHPYVPLGAVDPGLGYNVPALYSSDPAAREAREA--- 932  
Qy 1616 DYTTSQOMQVNLRPDVARGLSPRE--QPL--GLPYPATRGIIDLTNNPPTILVPHPGGSTP 1673  
Db 933 ---RERDLRNLKPGFE--VKPSELEPLHGVGP-----GLDP---PFRHGGGLAQ 975  
Qy 1674 PMDRITYPGTQITTPPRPYNSMSFGHPHTLAAAAAERE-----REREREKERE 1728



```

QY 1736 LYLRPGSEOP-----GRPGSHGYVRSPSVRTOETML-----QORPSVFOG----- 1777
Db 1453 QSLNPASFSLATQAKP-----EKSTLQRPQETVIRELQPOQOPTIERRDLOYITIS 1506
QY 1778 -----TNGTSVITP-----LDPTAQLRIMPLPAGGPSISQGL-----PASRYNTAA 1818
Db 1507 KEELSSGDS-LSPDPWKRDRAREKLEKQQMHIVDM-----LSKEIHELQNKGDRTAES 1559
QY 1819 DALAALVDAASAFQMDVSKTESKHE-----AARLEENLRSAAYSEQOQL 1866
Db 1560 DRLRKLMEWQFQRLQESKQKDEDEDDDDVTMLIMORLEARRAR--LQDEERRR 1617
QY 1867 EOKTLEVEKRSVQCLYTSSAPPSGKPPHSSVYSEAGKDKGPPPKSRYEELTRGKTT 1926
Db 1618 QOQLEEMKREVE-----DRVROEDG-----RHQEEERVK----- 1648
QY 1927 ITAANFIDVITITQIASDKDARERGSSSSSLSHRYETPSDA-----IEVISPA-SS 1981
Db 1649 -----RDAEKKRQEEGYYSRLEAERRRQHEEAARRLLEPEEPGLSR 1690
QY 1982 PAPPOEKLQTYQPEVVKANQAEENDPTROYEGPLHHYRPOQESPSPO-----QQLPPS 2033
Db 1691 PPLPQD-----YEPQSOSAPSPAPPPPPQORNASYLKTQVLSPD 1728
QY 2034 SOAEGMGQVPRTHRLITLADHICQIITQDFARNQVSSQTPPOOPTSTFQNSPSALVSTPV 2093
Db 1729 SLF-----TAKFVAYDD-----DDEENYV-----PAGP-----NSYSGSAGTTA 1763
QY 2094 RTKTSNRYSPESQAQSVHHQRPGRSVSPENLVDKSR 2129
Db 1764 GTYDAPRDTREKLSKSDADLPGSSGAPENLTFRRER 1799

```

Search completed: September 8, 2001, 14:42:40  
Job time: 447 sec